Human Behavior Profiling for Health Assessment Using Unsupervised Methods

Von der Carl-Friedrich-Gauß-Fakultät
der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines
Doktors Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Ju Wang
geboren am 28. August 1984
in Shandong, China

Eingereicht am:	01. 08. 2013
Disputation am:	11. 11. 2013
1. Referent:	Prof. Dr. Reinhold Haux
2. Referent:	Prof. Dr. Andreas Hein

2013

Acknowledgments

This thesis would not have been possible without many people's support.

First and foremost, I wish to express my sincere gratitude to my advisor, Prof. Reinhold Haux. He gave me the chance to do my doctoral studying in Peter L. Reichertz Institute for Medical Informatics. Prof. Haux is one of the nicest people I know. During the processing of my abroad studying, not only has Prof. Haux given me much advice with patience on my research, but also taught me how to manage myself. I would like to thank my second advisor Dr. Klaus-Hendrik Wolf. Klaus is always supportive. He has given me a lot insightful discussion and valuable advice on both academic and personal level. I am extremely grateful for his support.

Many thanks to the members of my committee, Prof. Lars Wolf and Prof. Andreas Hein. Thanks for their helpful advice and insightful comments.

I want to acknowledge the financial support from Chinese Scholarship Council (CSC) for three years funding. And I also want to convey my appreciate to the Office of Educational Affairs of the Embassy of P.R.China in Germany, thanks for their kindly and in time help.

I would like to express my appreciate to the GAL-NATARS study and all colleagues who work on it. Particularly I want to mention the study organizer Prof. Michael Marschollek, and my colleagues in Hannover Medical School. They always fulfilled my request. Without their help, I couldn't finish my work.

I am also indebted to my colleagues who have supported me during my stay in Braunschweig. Ute Zeisberg, Bianying Song and Dr. Markus Wagner, I thank their considerate help dealing with affairs during my daily life. Dr. Martin Kohlmann and Matthias Gietzelt have given me much useful advice on my thesis. Special thanks also to Prof. Friedrich M. Wahl and Prof. Jens-Peter Kreiß, they have given me very valuable advice on refining my work.

During my stay in Braunschweig, I have harvested precious friendships. Many thanks to all dear friends for their accompany, care, and encouragement. It was really happy to be with you. Many thanks to my friend Luke ten Kortenaar from Canada for his patience to help me refine this thesis.

Finally but not least, I really feel indebted to my dear parents, Xiufeng Wang and Yuexiang Xu, and my family, they are always behind me with their understanding and supportive attitude.

Thanks a lot!

Abstract

Human behavior profiling is an important approach for describing and assessing a patient's health status and rehabilitation conditions. In the context of ambient assisted living and the smart home, implementation of human behavior profiling is expected to occur through the use of pervasive computing, due to its capability of providing continuous observation. As for extraction of information from measured data, one typical way, referring to supervised methods, is to model the subject through training data and then to recognize current behavior or predict its future trend. However, considering the acceptance of users, due to the low adaptivity and high dependency on lab-setting, and the necessity of data labeling and model training, these types of methods are limited in practice to human behavior profiling in real-life scenarios.

Therefore, simple and unobtrusive sensors are relied upon to obtain daily behavior information. In spite of the incomplete observation, these sensors are able to provide key information, such as the user's position and the objects interacted with. Thus, unsupervised methods have to be designed based on this measurement. In contrast to supervised data analysis, unsupervised methods have inherent advantages: Firstly, data labeling and training are not necessary, which promotes to carrying out assistance in the real-life scenario. Secondly, they are more adaptive, making them suitable for use by different individuals. Thirdly, unknown knowledge might be discovered.

In order to propose unsupervised methods for human behavior profiling that can be practically applied, the following research is conducted in this doctoral thesis: First, abstractions of events and patterns of in-home behavior scenario are defined. Second, the discovering algorithm is derived, whereby regularly occurring sensor events that can represent lifestyle patterns can be discovered. Third, with the lifestyle depicted, the change of human behavior is modeled to present the variance of lifestyle. Aiming to investigate the effectiveness of these methods, they are applied to the datasets obtained in GAL-NATARS study, which is carried out in the setting of real-life within the framework of the Lower Saxony Research Network Design of Environments for Aging (GAL), and their effectiveness is evaluated through comparison with medical assessment results.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

In past decades, advances in technology have deeply affected approaches of diagnosis. The approaches to inspect patient's health status have broadened significantly. For instance, in the domain of medical imaging processing, a number of technologies, from the X-ray used for skeleton examination, computed tomography (CT), to Nuclear Magnetic Resonance Imaging (NMRI) and ultrasonography, have been broadly accepted by physicians and patients because of their remarkable advantages. Obviously, the interaction between supply and demand has never stopped. Humanity is experiencing what has never happened before, where both challenges and opportunities have arisen with the progress of the society. Among all of these, demographic structure change, which is often-cited, has attracted much attention and has been frequently reported, both nationally and internationally. By 2040, it is estimated that there will be 1.3 billion old people over the age of 65, which accounts for 14 percent of the total [1]. We have to admit, to some extent, that this has resulted from improved nutrition and healthcare conditions. However, as a result of the benefits that result from these improvements, there are also a number of arising issues. On the one hand, this implies an intensively growing costs of healthcare in that the elderly lack of independence and require more care, implying that more and more resources must be allocated to them, in terms of both funding and caregivers [2]. On the other hand, the unbalanced population proportion will result in more financial pressures. It is estimated there will be a lower proportion of young people to support the elderly in the future. According to the report released by United Nations, the potential support ratio (PSR), defined as the number of people aged 15-64 years for each older person aged 65 years or over, has declined from 12 in 1951 to 9 in 2009, and is expected to reach 4 in 2050 [3]. All in all, this evidence highlights the necessity to develop new methodologies in order to meet the upcoming challenges.

“The most profound technologies are those that disappear.” stated Weiser, the father of pervasive computing, over twenty years ago [4]. By integrating information processing into everyday objects and activities, pervasive computing aims to create ambient intelligence where networked devices embedded in the environment provide unobtrusive, continual, and reliable connectivity and also perform value-added services [5]. With pervasive computing, human experience and quality of life can be improved without explicit awareness of the underlying communication and computing technologies. The recent advances in ICT (Information and Communication Technology) have made it more practical and achievable to build pervasive environments with sensitivity and perception. Efficient tools and hardware platforms can provide reliable service. At present, more efficient products are available on the market and the cost reduction in both sensor and computing makes it easier to establish a system with pervasive computing. Firstly, various sensing devices are available so that more parameters, in terms of both wearable on body and embedded in living environment, can be measured. Secondly, the information can be acquired invisibly due to the miniaturization of sensors. Thirdly, by utilizing wireless communication, the system can be easily installed. Additionally, it is possible to process complex datasets, e.g. massive dataset and data with uncertainty, through the use of intelligent data analysis, such as machine learning, data mining, and artificial intelligence.

In the context of pervasive computing, with the aim of supporting healthcare, the concept of pervasive healthcare has emerged. Pervasive healthcare is defined in two aspects, which consists of the application of pervasive computing and making healthcare available everywhere, anytime-pervasively [6]. Pervasive computing is predicted to improve traditional healthcare in that ubiquitous and unobtrusive analytical, diagnostic and supportive information and documentary functions can be implemented [7]. Some capabilities, such as remote, automated and continuous patient monitoring and diagnosis, might make pervasive computing a tool advancing the shift towards new healthcare [7]. Through the sensors that are wearable or embedded in the living environment, many physical parameters such as heart rate, ECG, blood pressure, body temperature, respiratory activity and movements, as well as environmental parameters such as temperature and humidity, can be measured in real-time. This could profoundly influence the monitoring and management of chronic disease. Furthermore, the communication between patients, physicians, and other healthcare workers will be improved with well-established ubiquitous networks, enhancing the delivery of accurate medical information any-time and anywhere. For patients who do not require intensive care in hospital, but are still subject to relapse or other complications, it is possible to observe their recovery progress with a personal area network under out-of-hospital conditions, and necessary notifications can be transferred to corresponding personnel in case of emergency or a critical change in health status. The scenario information, including medical and environmental data can be obtained as well, which is helpful for diagnosis and devising of treatment plan [8]. Pervasive healthcare will contribute to the paradigm shift from the established centralized healthcare model to a pervasive, user-centered and preventive overall lifestyle health management [9]. Driven by quality and cost issues, there is a trend to move from managing illness to maintaining wellness [10], and to change healthcare systems from the current healthcare professional-centric system to a distributed network healthcare system whereby the individual becomes more active in the care process [9, 11]. Patient-centered healthcare also enables the involvement of the patients, thus enhancing patients' consciousness of keeping healthy lifestyle and obeying their physicians' suggestion.

To date, there have been considerable projects over the world focusing on the implementation of pervasive healthcare, but here only the most famous and established ones are mentioned so as to give a first impression of the current state of the art. House_n is a live-in laboratory developed in MIT for the study of ubiquitous technologies in home settings. Up to 30 sensors were applied to construct a network to obtain the information about objects that are manipulated, environmental conditions, and the use of appliances [12]. TigerPlace, a project of the University of Missouri, also used a various of sensors, such as bed sensor, motion sensor, cameras, etc., to detect human behavior [13]. CASAS, a multi-disciplinary research project at Washington State University, viewed the smart home as an intelligent agent that can perceive its environment through the use of sensors and can act upon the environment through the use of actuators [14, 15]. Aware Home at Georgia Institute of Technology is an interdisciplinary research project focusing on the health and well-being of aging adults, investigating how new technologies can impact the living at home [16, 17]. MobiHealth developed a body area network (BAN) to measure vital signs, including blood pressure, heart rate and ECG [18].

1.2. Motivation

Human behavior profiling is an approach to represent information related to human behavioral characteristics. It is able to outline an individual's or community's activity performance. This approach has been applied across a range of subjects such as marketing, criminology, and medicine. With human behavior profiling, critical characteristics of targets with respect to different applications are briefly and clearly described, whereby experts can quickly access the necessary information and make appropriate, informed decisions. With respect to medical applications, human behavior profiling is also an important approach for describing and assessing a patient's health and rehabilitation status. One typical application, for instance, is to profile ADLs (activities of daily living) associated with daily self-care activities within an individual's or community's residents.

Having in mind the background of the potential capabilities of pervasive healthcare, and the promising trend to shift to individual healthcare, it is expected that human behavior profiling will be implemented in a novel manner in the context of ambient assisted living and the smart home. Pervasive healthcare has laid the foundation for new approaches to human behavior profiling. With pervasive healthcare, it is much easier to measure parameters in real-time, which was difficult in the past. A range of behavioral records can be conducted automatically without manual involvement. For example, human activities can be recognized through sensors, such as wearable accelerometers, cameras and other motion detecting sensors deployed in living environment. As a result, an integrated profile can be developed, which supplements approaches to profiling.

There has been a range of research on the analysis of health-related human movements, in which the activity recognition is usually regarded as the main objective. Based on the application of various sensors, a lot of ongoing work focuses on detecting or identifying specific activities, such as fall detection, gait analysis, and statement transition from standing to sitting [19, 20, 21]. Considering knowledge extraction, intelligent data processing methods are also the subject of considerable research. With data mining and machine learning, classification and abnormal situation detection have been implemented. The data processing procedure can be summarized as firstly determining the model using a training dataset, and then applying the model to classify new dataset. However, some research issues are still open for applying new techniques to clinic. There is little research under real-life conditions due to some challenging factors, such as volunteer recruitment and cooperation between institutions. Furthermore, longitudinal studies are needed, which requires the evaluation of long term changes of some critical specific activities.

Considering the extraction of useful information from the recorded data, supervised methods are often used. Supervised methods model the target through training data and then recognize new states or predict future trends. However, there are some limitations to practical application of human behavior profiling within these methods. With supervised methods, data labeling is likely to be inaccurate. This is largely due to the following reasons: Firstly, variability exists across and within subjects [22, 23]. Human performance varies between different individuals, and it even varies under different health conditions of the same individual. Secondly, human activities are usually complex and concurrent. In daily life, several activities are often conducted in parallel, which makes it difficult to resolve these situations. Considering behavior data obtained in real-life, it is generally very impractical to label this data. There are three main ways to label behavior data. One is manually done by experimenter, another one is asking the subjects to label each activity they perform, and the third approach is letting the users or patients perform activities in a predefined order [24]. However, all these ways are time-consuming and unacceptable for users in real-life situations. In addition, users probably cannot conduct activities in a natural way and thus the measured data may not reflect reality. Furthermore, research into the application of these methods has been conducted primarily in laboratory environments, but not in real-life situations, and few studies did long term monitoring. Finally, there is limited research validating the effect on clinical outcome, which means there is no comparison with medical data.

Considering the practical application, unsupervised methods must be relied upon to support solutions. In contrast to supervised methods, unsupervised methods offer inherent advantages, which will address the previously discussed limitations of supervised methods [25]. The most advantage to supervised method is that data labeling and training are not necessary, which improves their applicability to real-life situations. Since the independent to training, they are more adaptive and suitable to be used by different individuals. Additionally, unknown knowledge might be discovered and useful features for categorization may also be found through the use of unsupervised methods. So far, there have been a range of research projects exploring the application of unobtrusive sensors and unsupervised methods, such as the CAR (circadian activity rhythms) models proposed to assess everyday life behavioral rhythms for the older generation [26] and the remote follow up through the monitoring of electrical activities on the residential power line [27].

With the goal of investigating the technical feasibility of the application of health enabling technologies to support independent living at home and acceptance of health enabling technology by users, GAL-NATARS Study (which stands for "Nutzung von assistierenden Technologien für das Assessment von Risikoprofilen bei Patienten mit stattgehabter Schenkelhalsfraktur"), within the framework of the Lower Saxony Research Network Design of Environments for Aging (GAL), has been conducted in the setting of real-life. The patients' real living environments were observed in this study. Only simple and unobtrusive sensors were used to detect patient's daily behaviors. Also there is no labeling work in this observation procedure. In addition, essential medical assessments were conducted periodically, which can be used to evaluate the effectiveness of the observation system. Consequently, the results of NATARS study are suitable material for investigating unsupervised methods.

1.3. Objectives

Considering the background information and motivation provided in the preceding paragraphs, the main objective and specific objectives of this thesis are stated in this section.

Based on the in-home healthcare information system, with the prospect of assisting the independent and self-sufficient life of elderly individuals with limited mobility or patients who are subject to relapse or in need of rehabilitation, the main objective of this thesis is to propose unsupervised methods for human behavior profiling, with the goal of being able to contribute to fully investigating a patient's health condition.

To achieve the main objective, the following sub-objectives of the current work must be accomplished.

- O1. Behavior pattern discovery
  In the setting of the in-home health monitoring system, the resident's activities trigger sensors so that the behavior can be manifested by the data collected by this system. For the purpose of representing an individual's behavior, the patterns implicated in the data have to be discovered.
- O2. Human behavior profiling
  Once the patterns are discovered, they can be used as the attributes of resident's behavior.
- O3. Analysis of behavioral change
  As an indicator of showing human health status, the behavior change is interesting for clinicians. By analyzing the correlation between sensor data and medical assessment data, it is expected that the profile change is able to predict the trend of human health.

1.4. Questions

To implement the objectives, the following questions must be considered and addressed.

- Q1. How can living environments be formally described?
- Q2. What kind of data preprocessing must be conducted?
- Q3. How can behavior patterns be extracted from datasets?
- Q4. What kind of metrics, which are able to reflect the changing health status of the patient in various contexts, can be selected and defined?
- Q5. How can the relation between the behavior pattern obtained through the proposed methods and the results obtained through medical assessments be evaluated? Furthermore, how can the effectiveness of the proposed methods be assessed?

1.5. Outline of the Thesis

The remaining chapters of this thesis are organized with the following structure:

Chapter 2, titled with *foundation and materials*, firstly review the in-home health monitoring. Following content covers foundations of two-side, including medical background and mathematical background. At last, the GAl-NATARS Study is introduced. This study provides the data source used in this thesis.

Chapter 3 abstracts the living environment. The architecture is mapped onto a graph named as *@home graph*. Two kinds of events are formalized to denote the interaction between resident and objects. The last section describes event patterns based on the assumption that periodic behaviors are usually carried out.

Chapter 4 starts with dealing with heterogeneous data. Two kinds of are extracted from raw datasets. Afterwards, 4 proposes the scheme that is able to discover behavior patterns. Based on extracted events and behavior patterns, a number of metrics are defined to profile human behavior.

By applying the theories and methods to the datasets collected in GAL-NATARS, chapter 5 presents the results of four subjects.

Chapter 6 discusses the whole work from two aspects in detail. Firstly, discussing the achievement of objectives of this thesis. Secondly, discussing the questions defined this in this chapter.

Finally, chapter 7 concludes the whole work, and comments on the future work on this topic as well.

Chapter 2

Foundation and Materials

As mentioned in the introduction, the relationship between application area and technical progressing is actually the interaction between demand and supply. With respect to medical informatics with interdisciplinary characteristics, covering medicine and informatics, both of these aspects should be taken into account so as to conduct significant research.

In order to implement the objective presented in the previous chapter, the backgrounds from three aspects are necessary, i.e., in-home health monitoring, medical background, and mathematical theories. Before implementing the objective, this chapter will firstly cover these backgrounds. Section 2.1 gives a review of the background of the monitoring during everyday life, the in-home health monitoring system, which is a realization form of the human behavior record; In section 2.2, some medical backgrounds that is about medical assessments is briefly introduced, which confines the design criteria of the current work; some related mathematical theories will be introduced in section 2.3, and the essential methods for data processing and analysis are included. At last, as the materials providing data source, a real field study, the GAL-NATARS study is introduced.

2.1. In-home Health Monitoring

The current demographic change in aging population, with the impacts on the healthcare system, is demanding changes in healthcare solutions to provide efficient service. The world's potential support ratio (PSR), which is defined as the number of persons aged 15 to 64 per every person aged 65 years or over, has declined from 12 in 1951 to 9 in 2009, and is expected to reach 4 in 2050 [3], and the trend is happening in both developed and developing countries. This means that there will be fewer people to support the living of the elderly. What makes the situation worse is the declining of the amount of workforce for healthcare. Meanwhile, a dramatic increase in the prevalence of chronic diseases, disability, frailty among the elderly community is expected since increasing life expectancy [28]. As a result, it is necessary to offer continuous healthcare, monitoring and management for the well-being of the elderly, which are able to reduce the human intervention [28]. The emerging of pervasive healthcare is hoped to realize this target in brand novel methods.

Pervasive healthcare is promising in that it may benefit many aspects. The patient benefits from pervasive in-home health monitoring. Since the advances of techniques, unobtrusive and wearable sensors are able to collect vast amounts of data for next-generation clinical trials [29]. The remote monitoring capability allows to interpret measured data and provide intervention in time [28, 30]. Early detection can be provided for better management of their living, so as to prevent accidents and hospitalizations. These capabilities are suitable for controlling chronic diseases, especially for the elderly to remain independent living. Healthcare providers can also benefit from this kind of monitoring system. It is possible to provide efficient management of resource allocation with better informed care planning and coordination [28]. And better quality of care can be provided with early preparation. Considering the less care workforces, the healthcare providers can profit from reducing the burden placed on caregivers.

Technological maturity has been promoting the development of the pervasive in-home health monitoring. During past decades, the advances in semiconductor, sensing technology, computing have made it more practical to implement new healthcare approaches. As a realization of pervasive healthcare, in-home health monitoring system can improve the quality of life for the elderly, including independent living, health status tracing and treatment management. It has been proved that care may be provided more efficiently in home rather than in hospital [31]. Consequently, in-home health monitoring is expected to meet the demands of the current demographic change. In this section a review of state-of-art of in-home health monitoring is given. Note that the term *in-home health monitoring* refers to the monitoring system that can automatically measure parameters and assess the resident's health status at home. Here the forms can be intelligent living system, wearable system or their integration.

2.1.1. Definition of In-home Health Monitoring

Towards a novel way of healthcare, a number of concepts related to in-home health monitoring have been proposed. Demiris stated that "*smart home* is a residency setting equipped with a set of advanced electronics and automated devices designed for care delivery, remote monitoring, early detection of problems or emergency cases and maximization of patient safety" [32]. Similarly, Chen commented that this term refers to residence equipped with technology that allows to monitor its inhabitants and encourage independence and the maintenance of good health [31]. The term "Telemedicine" literally means "healing at a distance". According to the report from WHO, it is stated as "the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities" [33]. From another aspect, telemedicine can also be defined as "the use of audio, video, and other telecommunications and electronic information processing technologies for the transmission of information and data relevant to the diagnosis and treatment of medical conditions, or to provide health services or aid health care personnel at distant sites" [34]. "Health-enabling technologies" are a term referring to information and communication technologies for creating sustainable conditions for self-sufficient and self-determined lifestyle [35], and sensor-enhanced health information systems play a major role in this context [36].

On account of these overlapped and coincident concepts, the definition of in-home health monitoring can be defined as: Characterized by pervasive and individualized health care, in-home health monitoring is the health service for the resident, which is implemented in daily living environment based on telecommunication, information and sensor enhanced environment, and can be integrated in conventional healthcare system.

2.1.2. The Structure of In-home Health Monitoring

With respect to the structure, some summaries have been made. Haux et al. summarized four contemporary paradigms for the ICT (Information and Communication Technologies) architecture for health-enabling technologies [35], and three of them are closely related to the in-home health monitoring, i.e. Person-centered ICT architecture (PCA), Home-centered ICT architecture (HCA), and Telehealth service-centered ICT architecture (TCA). According to Haux et al., *new ways of living* – with the aim of supporting self-determined and self-sufficient life, and *new ways of health care* – with a strong focus of outpatient care and a significant potential for prevention – are expected in the future [35]. In order to demonstrate the interweaving structure, a diagram with two-circle has been depicted (see Figure 2.1). Data are measured from patient and service is offered based on data analysis. In both circles, the patient is the center of the service.

Even though there have been a range of projects aiming to explore in-home monitoring systems on different specific applications, in general they are mainly in common within the existing in-home health monitoring systems with respect to the structure.

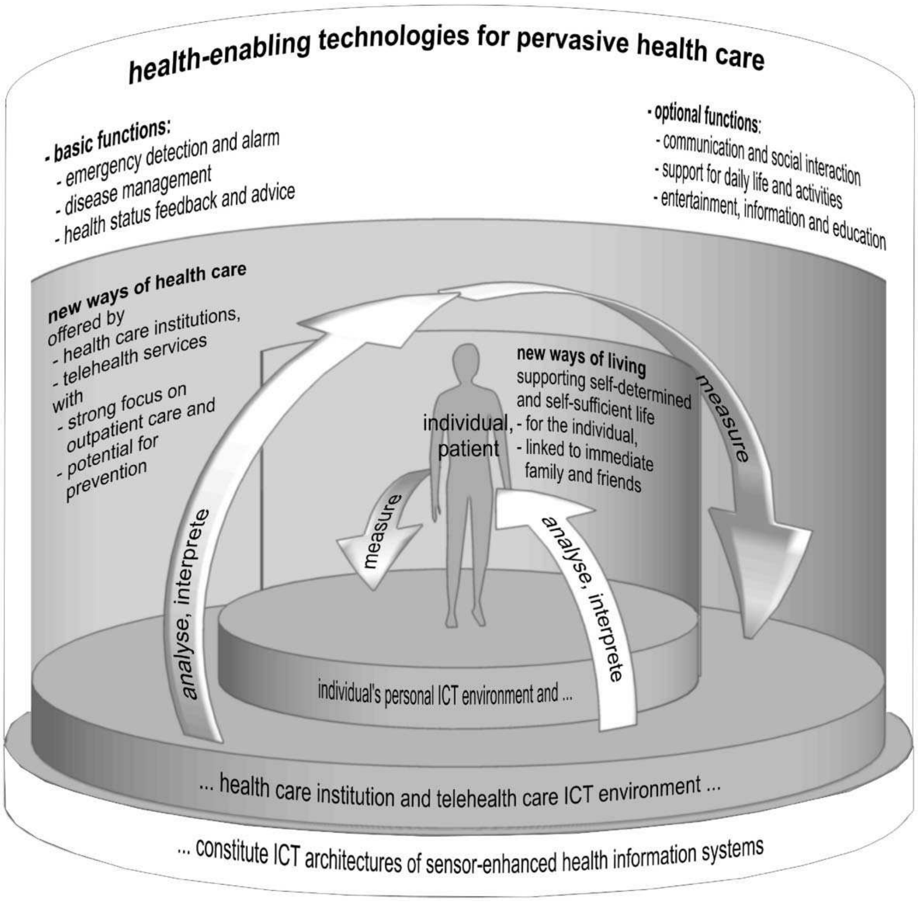


Figure 2.1.: Pervasive health care with health-enabling technologies viewed as a double-circle [35].

From the aspect of information flow, the system can be described as following: Sensors are deployed in the daily living environment or worn on human body, constructing personal area networks (PAN) or body area networks (BAN) wired or wireless. The data gathered through sensors is stored in the base station or the gateway, e.g., PDA (personal digital assistant) or personal computer, and then transmitted to the remote server through wide network. Information extracting is conducted through data analysis methods either in local terminal or in remote terminal. As a result, clinicians or caregivers on the remote side can access the information so as to carry out corresponding actions. As feedback, the instruction or alert is delivered to the patient through short message service (SMS), E-mail, telephone or teleconference system. The process is illustrated in Figure 2.2.

For the prospect of the application of in-home monitoring, an forecast view of a future scenario for health care environment is depicted in a survey, which is also in line with the description above [30]. In the future scenarios, resident, community

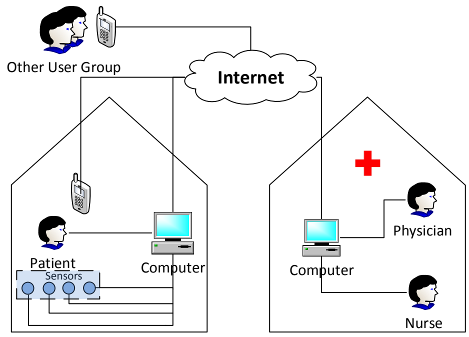


Figure 2.2.: The structure of in-home health monitoring

and professional institution are connected into a whole system. Living environment is no longer separated from healthcare. All available information is collected through embedded sensors and the information flow is transmitted between involved users.

2.1.2.1. Sensors and Parameters

Because of the advance in semiconductor, various kinds of sensors are available to measure a range of parameters. The categorization of sensors is presented in Figure 2.3 [37]. With these sensors, usually a range of health related parameters (see Figure 2.4) can be measured through in-home health monitoring.

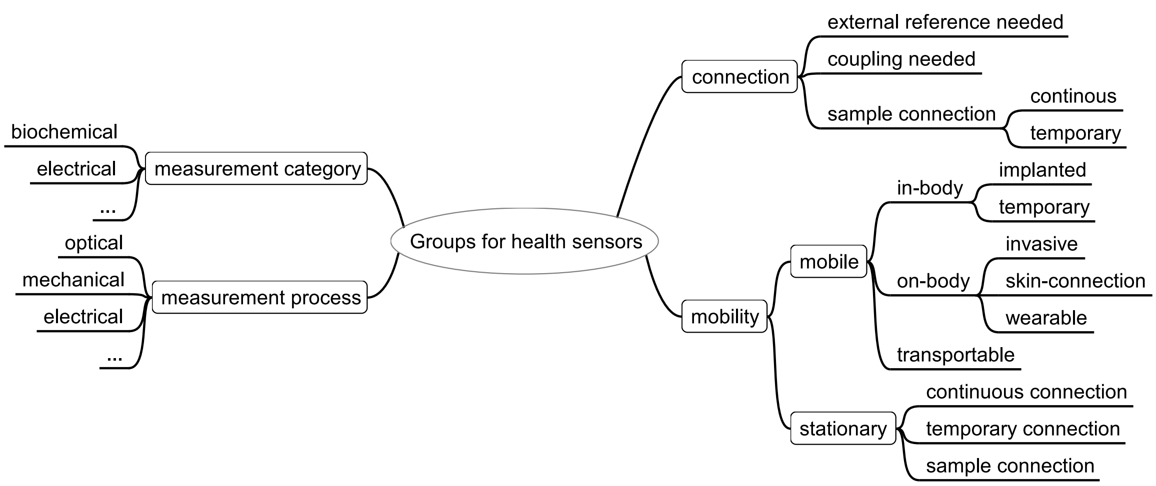


Figure 2.3.: Categorization of sensors [37]

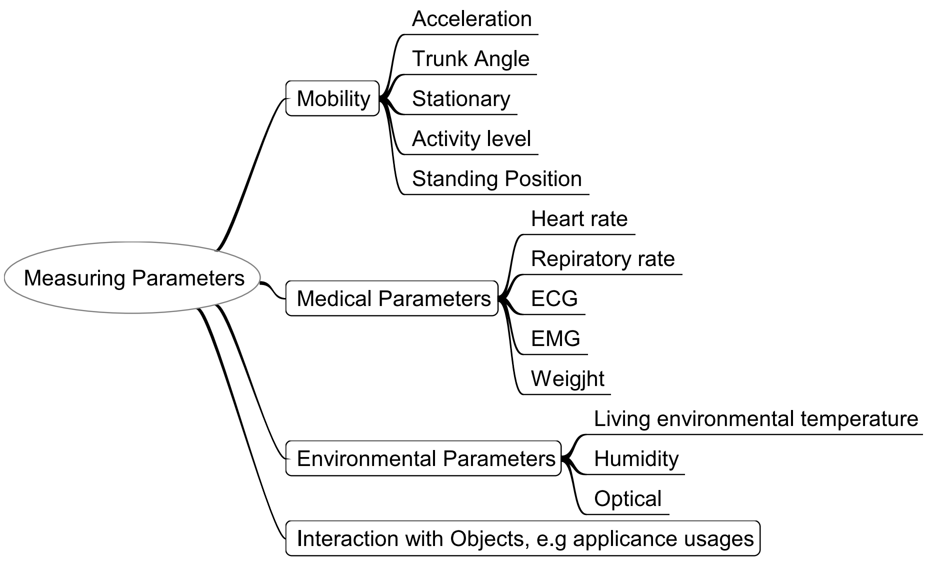


Figure 2.4.: Measuring parameter of in-home healthcare

2.1.2.2. Users

With respect to applications, users are the people who need assistance, and those people who provide assistance. According to [31], the users for smart homes are those people with disability, either mentally or physically and those persons who provide care for the people needing assistance. Similarly, here we summarize the users of in-home monitoring as following:

- The older people living alone;
- The disabled people who suffer from cognitive and/or physical impairment;
- Nurses and clinicians.

2.1.3. Applications

Considering users' backgrounds, especially those who need assistance, a variety of systems and projects have focused on exploring the potential of the in-home monitoring to assist the healthcare process.

Because of the potential of continuous and real-time monitoring, some research projects aimed to provide assistance to some specific diseases management, such as, the devices developed for patients with diabetes [38], the remote monitoring system for heart failure [39]. In these research, the vital signs, including temperature, blood pressure, pulse, respiratory rate, are always of great interest to both the physicians and the caregivers. Through wearable system, it is possible to automatically measure vital signs continuously or periodically. For instance, a Body Area Network (BAN) was developed to transfer ECG signal, respiratory rate, and blood pressure [40]; in [41] an in-ear measuring system was proposed for heart function assessment. Activity recognition is also a hot research point. Through activity recognition, some specific situations can be detected, such as the imbalanced diet detection [42] and the Parkinson's disease [43]. The information gained via activity recognition can also be used for higher level behavior analysis. Not only through the direct detection of human parameters, but also the environment is able to provide helpful information. For instance, in the setting of smart home, Fleury et al. tried to detect distress situations and deduce the activity of daily living performed through gathering sound signals that are generated within the living environment [44].

Among the applications, those for the older people have the most potential. Because the functional status is by far the most important factor affecting quality of life and health care utilization in old age, valid outcome measures of physical activity and physical functioning in aged individuals are of utmost importance [45]. Long-term tracing refers to continuous monitoring of one or multi-parameter,persisting several months or even several years. Long-term tracing can be classified as health status prediction and lifestyle evaluation. Based on the application of inertia sensors, camera, presence sensors, the activity recognition and the behavior modeling have permitted to establish motion-based behavior of the elderly upon their habits in terms of displacement and activity level, leading to the implementation of long-term tracing [26]. Tâm [46] recognized daily routines as a probabilistic combination of activity patterns, where topic models are used and a unsupervised method was also discussed in his work.

Though various approaches are used, varying in term of both sensors and structures, are used, they can be assorted into several categories. There have been some summaries of the application of smart home, telemedicine, telehealthcare etc. For instance, from the aspect of application types, a conclusion was made in Mathie's dissertation [47], consisting of Tele-consultation systems, Patient tele-Monitoring systems, Personal emergency response systems, Medication dispensing systems, and unobtrusive monitoring systems. Demiris et al. sorted the health-related smart home technologies into six categories, i.e. physiological monitoring, safety monitoring and assistance, security monitoring and assistance, social interaction, monitoring and assistance, and cognitive and sensory assistance [48]. A similar result was also presented in [35], in which they are specified into basic and optional applications.

Basic functions for pervasive healthcare comprise:

- B1. Emergency detection and alarm
- B2. Disease management (for chronic diseases)

- B3. Health status feedback and advice (consultation)

And optional functions:

- O1. Communication and social interaction
- O2. Support for daily life and activities
- O3. Entertainment, information and education

According to the categorization of [49], the application can be demonstrated through a mind map (see Figure 2.5).

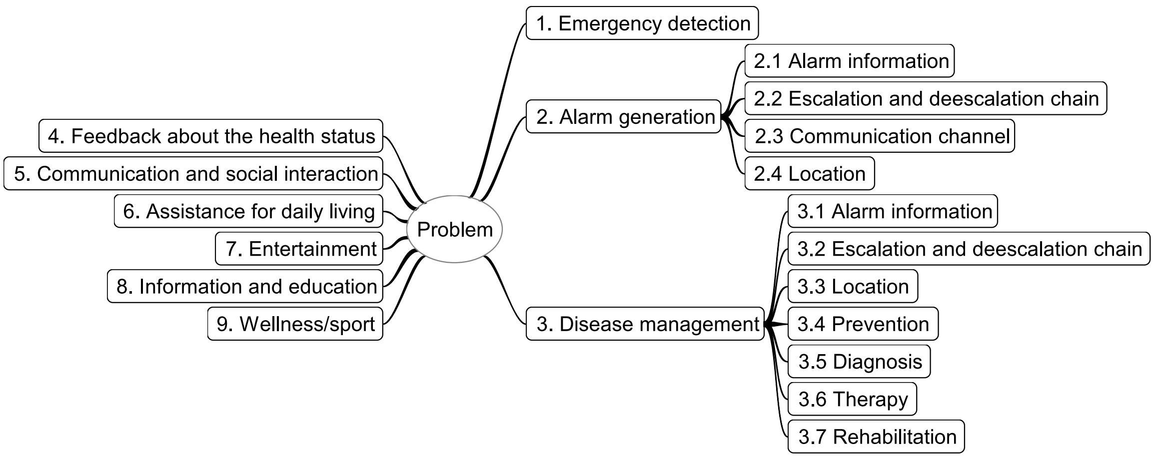


Figure 2.5.: Categorization of application of in-home monitoring [49]

2.1.4. Intelligent Data Analysis

Intelligent Data Analysis (IDA), which is process-centered and deals with the whole process of data analysis, is a research field that devotes methods to transform data into information by exploiting the available domain knowledge [50, 51]. The basic philosophy of medical informatics is to be application driven and its goal is to develop, adapt, or re-use existing methods to solve specific problems [50]. Therefore, towards better support for medicine, IDA has been broadly employed in medical application field.

In medical data analysis, the demand on dealing with data interpretation procedure intelligently lies in the following three aspects. Firstly, the explosion of data calls for novel data processing methods to meet the requirements. With the application of multi-measure approaches and information systems, the amount of data has been increasing dramatically in past decades. It is expected that this trend will continue in the future. Electronic data have become ubiquitous in biomedical field, thus resulting in a corresponding need for analytical methods [51]. In view of health information system, one the one hand, the regional healthcare information systems can provide more data and electronic health records (EHRs) can further expand to national or global scale [52]; on the other hand, with a background of pervasive computing, all the events affecting a person's health can be collected electronically throughout one's lifetime, the personal health record (PHR), or the personal life record (PLR) may expand the amount of medical data or data related to health as well [51]. Secondly, high cost of gathering data need IDA to deal with data with uncertainty. Some experiments or studies are costly due to personnel and instrumentation involvement and the patients' discomfort. The data may be affected by several sources of uncertainty, such as measurement errors, data missing, errors of data coding or errors arising when information is extracted from textural reports or medical charts. For the sake of making full use of the data, the IDA would be the key approach. Thirdly, some specific situations need to process a amount of data in time. For instance, real-time decision support based on ICU monitoring data requires processing the multidimensional data in limited time, and data fusion needs to be conducted for the interpretation of data from wearable sensors as the limitation of sensor's memory. Because of the data-rich feature of medicine, the sophisticated methods and approaches that are able to analyse the increasingly *big* and *complex* datasets are required.

According to Peek, the recurrent and emerging topics in intelligent data analysis for medical applications are systems biology and metabolic pathway modeling, gene expression data modeling, quality of care assessment, and signal processing from in-home monitoring systems [53]. This section will particularly focus on data processing for in-home healthcare system. Firstly the process of IDA is briefly described. Following is a review of some common methods for in-home healthcare monitoring. Finally, these lead to the conceiving of unsupervised in-home healthcare monitoring.

2.1.4.1. The Process of Intelligent Data Analysis

IDA is a very broad domain that integrated the existing methods and knowledge to develop fast, smart and flexible analysis methods with respect to specific problems. Statistical analysis, machine learning, and data mining are mainly employed in implementation. The core of current IDA methods are classification models, temporal abstraction and visualization. It is difficult to describe a common specific processing for IDA in that it involves various theories. Here only a brief summary is given. First of all, the objective should be identified with regard to the problem needing to be addressed. Secondly, based on the objective, selecting the appropriate data processing methods or theory. In this step, multiple methods may be used. Thirdly, adhering the process as to corresponding methods that are chosen. Finally, verifying the results and modifying the methods if necessary until obtaining the results expected (see Figure 2.6).

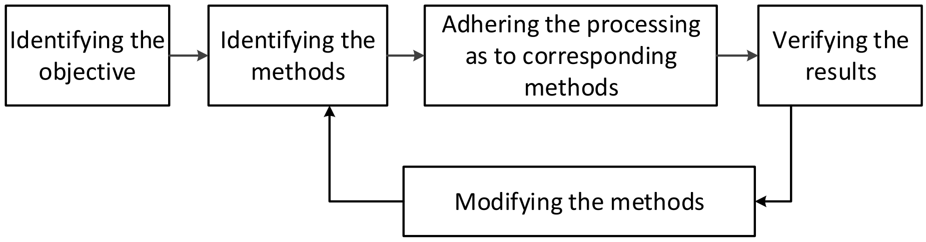


Figure 2.6.: **The processing of IDA**

2.1.4.2. Analysis Methods

Although there are various data sources from different types of sensors, the general data analysis procedure can be described as: at first, the candidate features are extracted, and feature dimensions should be reduced if necessary; following is the structure identification, including classification or clustering, and time series analysis for information discovery, etc. [49]; With regard to some specific applications, it is necessary to regress or predict the pattern trends. The category of these methods are demonstrated through a mind map in [49] (see Figure 2.7).

In this section, some frequently used methods for IDA and the applications in in-home health monitoring will be briefly reviewed. Meanwhile, the advantages and disadvantages corresponding to different methods are put forward.

2.1.4.3. Feature Set Generation

Mostly the raw data contains noise due to uncertainty generated during measuring and collecting, and they can not be used directly to reflect the character of the objects observed. Thus features must be generated and selected for desired structure identification methods. This section firstly reviews a few frequently used feature generation methods [54]. Following is the feature selection and dimension reduction.

Time-domain features

Time-domain features are the most intuitive features. These are derived directly from a window of sensor data and are typically statistical measures, i.e. mean, median, variance, skewness, kurtosis and interquartile range [54].

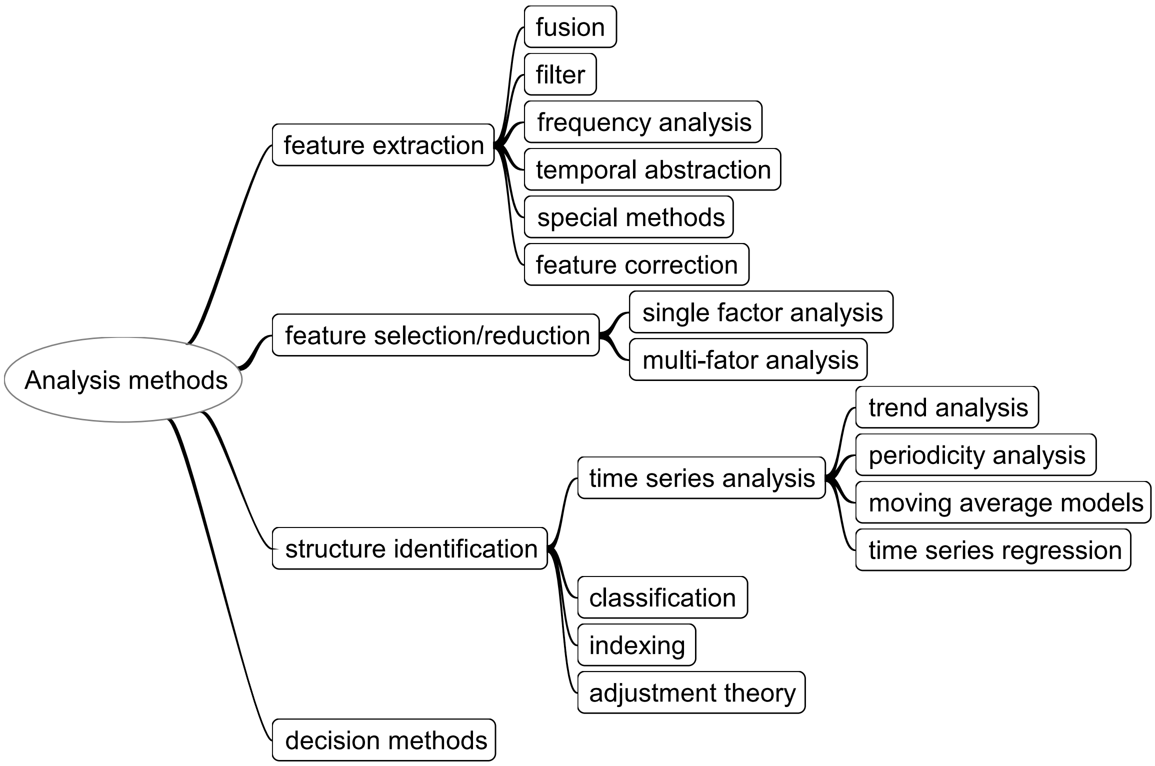


Figure 2.7.: Mind map of sensor analysis methods [49].

Frequency-domain features

With frequency domain transform, which is usually fast Fourier transform (FFT), frequency spectrum can be derived, which is able to intuitively reflect frequency distribution. Both the frequency components and their values can be employed to characterize data.

Time-frequency features

At times, it is not enough to apply either time-domain features or frequency-domain features. Wavelet analysis provides convenient tool to investigate both time and frequency characteristics of the signal. With this approach, the original signal is successively decomposed into separate low and high pass filtered signals, referred to as approximation and detail coefficients respectively. Because these coefficients characterize the original signal along its entire length, they contain information on temporal changes in frequency content. As a result, wavelet techniques can be used to analyse and characterize non-stationary signals that in which frequency context changes over time.

Feature selection and extraction methods

In many cases, some features derived are redundant and irrelevant. In order to improve the efficiency and reduce the complexity of further processing, the feature set that show little variation between repetitions of the same class and show considerable vary between different classes should be extracted [25].

For obtaining efficient feature sets, there are mainly two approaches, i.e. feature selection and feature extraction [55]. The first approach is to seek a subset of features out of the available set of features. The second one is to find a transformation from the higher measurements to a relatively lower dimensional feature spaces. Both of them need the optimisation of the criteria function, which is usually discrimination function. As to in-home healthcare (behavioral profiling), on the one hand, different individuals might perform the same activity in a variety of different ways. On the other hand, with respect to different individuals, the activities they performed might imply different meaning. This can lead to substantial variability in the feature sets derived from sensors in practical even for the same system.

2.1.4.4. Classification and Clustering

Numerous literatures have reviewed the data analysis methods, especially on the data classification and clustering [22, 54, 56, 57]. The often mentioned classifications include threshold-based classification, hierarchical methods, decision trees, k-nearest neighbors, Artificial Neural Networks (ANN), support vector machines (SVMs), Naïve Bayes networks, Markov chains, and hidden Markov models, etc.

Most classification methods are built on the base of the prior knowledge. Threshold-based classification is implemented by simply comparing the feature derived to a predefined threshold to determine whether a particular event occurs. This approach is suitable to detect the abnormal events, e.g. fall detection [19], [58]. Training data, which is labeled manually, is required to implement supervised methods, including hierarchical methods, k-nearest neighbors, decision tree, artificial neural networks, support vector machines, Markov chain and hidden Markov models. Hierarchical methods and decision tree need inspection to the training data to construct the decision node, while the former is conducted manually and the latter is conducted automatically by algorithm. K-nearest neighbors classifies objects based on the closest training data samples. Artificial Neural Networks (ANNs) [59] is a mathematical model used to model complex relationships between inputs and outputs. Its structure consists of three groups of artificial neurons, i.e. inputs, hidden and outputs. Sometimes the hidden layer can be multilayer (see Figure 2.8). In the field of in-home health care, the inputs are usually features derived from sensor data, and the outputs are the activity patterns, health status etc. To implement ANNs, as other classification methods, firstly the model must be trained, which means to obtain the weight on the synapses through training, and then unknown data is input to the model and the output is the labeled categories. Support vector machines (SVMs) [25] is also a pop-

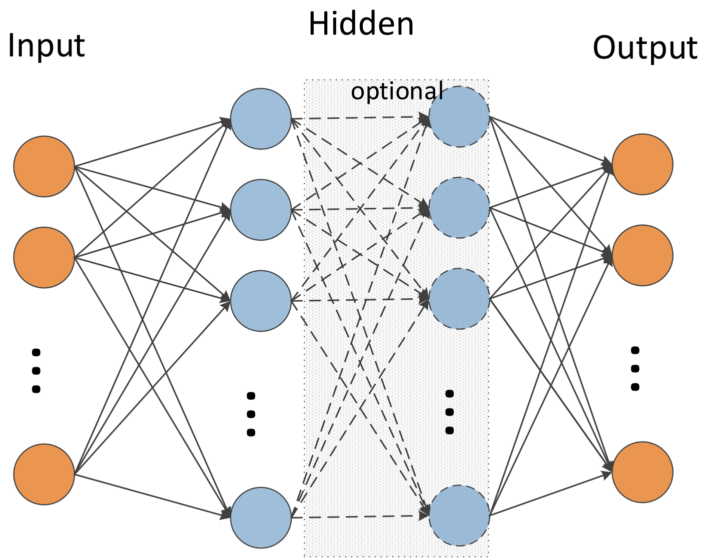


Figure 2.8.: The structure of artificial neural network with three layers.

ular machine learning method which is based on finding optimal separating decision hyperplanes between patterns with maximum margin. The preprocessing transforms data to higher dimension through kernel function, and then a linear separation in the new space becomes equivalent to a non-linear classification in the original space. Given the training data, the weight vector must be determined via SVM training.

Bayes networks, Markov chain, and hidden Markov model are based on the conditional probabilities and their parameters, which are state transition probabilities are also determined by the training process. In order to complement the classification, it is necessary to combine different classifiers in some cases. These methods have been used in a lot of applications. However, most of the applications are mainly implemented in laboratory environments, where the sensor data are measured under intensive observing, which makes it easier to implement data labeling. For the applications in real-life situations such as the in-home monitoring, there are dramatic variance between individuals. Two obvious issues make it impractical to apply supervised methods in real-life situations. One is inconvenience of data labeling. As the data are generated while the resident is carrying out activities, manually labeling data will bring additional burden on users. Another is the unreliable accuracy of the model. The fine trained model on one resident is usually not adaptable to the others. The training processing have to be repeat as to individuals. Consequently, it is impractical to be used directly in real-life in-home healthcare for most of these methods.

In contrast to supervised methods, unsupervised approaches are attracting more attention since the inherent advantages. With unsupervised methods, data labeling and model training are not necessary. So they are more adaptive and achievable for real life application. They are able to discover new knowledge and many data mining methods are based on unsupervised methods. K-means intends to partition n observation into k clusters, in which each observation belong to the cluster with the nearest mean. Hierarchical clustering builds a hierarchy of clusters. the data distribution is modeled via mixture models model. Unsupervised learning has the potential to be used as the first stage of a system for detecting adverse events.

2.1.4.5. Information Discovery Over Time

Both clinical management of patients and clinical research are essentially time-oriented endeavors [60]. Time is a factor that cannot be ignored in medicine. The change of symptom is closely related to time variation, and the procedure of treatment is the control of the change in the patient's health over time, which is represented by the change of measured parameters. With temporal analysis, the doctors and the care-givers can acquire valuable temporal information to treat the patients. For in-home healthcare monitoring (e.g., behavioral profiling), the events occur in time order, and distribute over time axis. The often-used sensors or systems are able to record the timestamps, which is able to provide time information to analyze the situations.

For the sake of discovering hidden information from time related data, the temporal data mining has been explored as an extension of data mining. With temporal data mining, the contextual relationship and the temporal proximity can be inferred, and in some cases , as well as the causality that is overlooked when the temporal component is regarded as simple numerical attribute [61].

The time related data usually refers to *time series*, which is a series of continuous, real-valued elements, and *time sequences*, which is a sequence of nominal symbols from a particular alphabet [62]. For processing time series, they are usually discretized and symbolized, so that they are converted into time sequences [63]. The tasks of temporal data mining include frequent episode discovery, association rules discovery, classification and clustering, and evolution and prediction. Frequent subsequences can convey the frequently occurring events, thus representing the performer's behavioral pattern. With episode discovery, the episodes that occur frequently can be derived. In some applications, it is used to characterize the data or to detect the abnormal situation. The events that accompanied occur constitute association rules. Temporal association rules might suggest a number of interpretations [61]:

- the earlier event plays some role in causing the later event,
- there is a third event that causes both other events,
- the confluence of events is coincidental.

Based on the results of feature extraction of the time sequence, the whole sequence can be split into a number of segments which can be further classified or clustered into different groups. The results of the frequent episodes discovery and association rules discovery can be used to characterize the sequences or series, and they have their own temporal properties. As a result, the classification and clustering may be implemented in a higher dimension which can be regarded as *pattern evolution* [61].

There have been some research that explored methods for temporal sequence processing. Chikhaoui et al. [64] proposed the ADR-SPLDA for activity discovery and recognition, whereby Apriori algorithm was firstly adapted on the input sequences and then its output was input to the topic model LDA(latent Dirichlet Allocation). Duchêne et al. [65] adopted the projection algorithm to the time series for feature extraction and clustering the series and the series are distinguished between normal and abnormal through clustering methods. Jakkula et al. used temporal data mining for anomaly detection [66].

2.1.4.6. Conceiving of Unsupervised Implementation

Having the background of in-home health monitoring above in mind, a scheme for in-home health monitoring in an unsupervised manner is conceived. As an example, consider the following scenario: An old person lives alone, and he mostly stays at home He has a regular lifestyle, which means he performs the activities in a rhythmic way. If a sensor-enhanced environment is built in his home, the sensors mounted on the furnitures, appliances, and his body can make corresponding records once he moves into the detecting area or uses the appliances and furnitures In this case, this old person is the trigger of his home and all the records are in temporal order. Based on the assumption that his lifestyle is rhythmic, there should be certain patterns reflected by the record. Aiming to analyze his behavior, the temporal data mining method can be used on the temporal dataset to extract his behavioral pattern. As for the trend prediction of the user's health status, the pattern evolution may be also necessary.

2.2. Medical Background

This section describes the background in medical aspect connecting to in-home healthcare and human behavior. Human health status is shown not only through inherent indicators such as vital signs and biomarkers, but also the external ones, such as activities and emotion. This section covers three kinds of common healthcare assessments, i.e., activities of daily living, lifestyle regularity and health-related quality of life.

2.2.1. Activities of Daily Living

For survival, normally humanity have to perform a variety of activities in everyday life, referring to activities of daily living (ADLs). In healthcare, this term comprises any daily activities concerning self-care, such as feeding, bathing, dressing, grooming, work, home-making, and leisure.

ADLs can be classified into two categories, i.e., basic ADLs (or personal ADLs) and instrumental ADLs [67]. As to basic ADLs (B-ADLs), they mainly refer to self-care tasks, including bowel and bladder control and management, dressing, feeding, transferring, personal hygiene, grooming, and sleeping. Obviously, maintaining the independence in self-care is essential for health and safety [68]. In contrast to B-ADLs, instrumental ADLs (I-ADLs) refer to more complex activities that are performed by a resident concerning community settings in a normal day, such as managing money, shopping, telephone usage, traveling in community, housekeeping, preparing meals, and taking medications. To accomplish these ADLs, a certain physical and mental ability are necessary. Increasing inability to perform either B-ADLs or I-ADLs may result in the need for care facility placement.

Both B-ADLs and I-ADLs can be seen as benchmark indicators of health and safety, and the loss of independence in the ability to personally perform ADLs is a predictor of nursing home admission, using of physician services, using of hospital services, and even mortality. Studies have shown that the functional dependence significantly predicted later institutionalization [69, 70]. Normal aging changes, accident, acute illness, worsening chronic illness, and hospitalization contribute to the decline in the ability to independently perform tasks of living in the community. Besides, depression and other psychiatric illness may also lead to functional ability [71]. Depending on situations or health status, the caused limitation may be temporary or permanent, and the patients should be promoted to the greatest degree of independence through treatment or rehabilitation.

Since the physical functional decline may be the first sign of a changing health status, both the ability and the inability to perform ADLs are usually adopted to assess the person's health status, especially regarding to the older people. In order to provide objective data on the patients health status, functional assessments are usually carried out by the healthcare professionals to detect problems in performing activities of daily living and to plan according healthcare. The assessment of ADLs has been applied to the patients with various of illnesses that can result in limitation of motility or cognitive, such as dementia, Alzheimer's disease, stroke, and neck fracture and fall-prone. What can be easily found is all of these illnesses are likely to cause increasing appearance of activity disorder or significant decline in activity level.

In order to evaluate and quantify the change in ADLs, some assessments have been developed. Katz Index of ADLs is the assessment designed for assessing B-ADLs [72]. A number of six functions are included in this instrument, i.e., bathing, dressing, toileting, transferring, continence, and feeding. Clients are scored yes or no for independence in each of the six functions, where the high score indicates higher independence. Katz Index is most effectively used for elderly population. Similarly, Barthel ADLs index comprises 10 items, which are similar to such as Katz Index. Specially, this index focus on recording what a patient does, rather than what a patient could do. This assessment can be used to determine a baseline level of functioning and monitor improvement in activities of daily living over time. Because I-ADLs function is usually lost before B-ADLs, the incipient physical or cognitive decline might be detected through I-ADLs assessment. Besides B-ADLs, I-ADLs are also taken into account by some assessment instruments. Bristol Activities of Daily Living Scale is designed to reveal the everyday ability of the people with memory difficulties of one form or another. A total of 20 activities are included, containing both B-ADLs and I-ADLs. And for each activity, there are five statements referring to different levels of ability. Another popular I-ADLs assessment is Lawton Instrumental Activities of Daily Living Scale (LIADLS). Competence in skills such as shopping, cooking, and managing finances are required for independent living in community. To assess more complex ADLs, Lawton et al. identified 8 items in LIADLS are scored regarding to different independence levels [73]. The 8 items are using of telephone, shopping, preparing food, housekeeping, doing laundry, using transportation, handling, medications, and handling finances. Whereas the low score on telephone, self-medicating, and managing finances may indicate the decline in cognitive functions in community-dwelling, the low score on housekeeping may more obviously point to the problems in physical function.

2.2.2. Lifestyle Regularity

Humanity perform certain activities during daily life and these daily activities are not absolutely disorganized regarding normal people. As known, some biological processes occurring inherent represent oscillation of about 24 hours, referring to *circadian rhythms*. Similarly, daily life is structured into daily patterns and cycles and these patterns of daily life can also be described with a similar term, *behavioral circadian rhythms* [74]. Whereas the inherent circadian rhythms are driven by endogenous circadian pacemaker (ECP), which are also called time cues or zeitgebers, the behavioral circadian rhythms are driven by the person's social and physical environment, regarding social contacts, meals, work, recreation, exercise and bedtimes [74]. For instance, due to difference in either inherent or behavioral circadian rhythms, people may exhibit morningness and eveningness [75]; due to living stages and health conditions, huge diversity on performance of maintaining a stable behavior circadian rhythm, resulting in the variance in lifestyle regularity.

In order to quantify the lifestyle regularity, Monk et. al. developed a diary-like instrument that is named as *Social Rhythm Metric* (SRM) [76]. In the SRM-17 version, a sum of 17 kinds of activities are included and their occurring time is recorded. The 17 activities are (1) Get out of bed, (2) First contact with another person, (3) Morning beverage, (4) Breakfast, (5) Go outside, (6) Start work, housework, or volunteer activities, (7) Lunch, (8) Afternoon nap, (9) Have dinner, (10) Exercise, (11) Evening snack or drink, (12) TV news, (13) Other TV program, (14) Activity A, (15) Activity B, (16) Return home last time, (17) Go to bed. Note that Activity A and Activity B depend on specific studies.

Through this assessment, studies have shown that lifestyle regularity is associated with human health status. Higher lifestyle regularity may prevent long-lasting bereavement-related depressive [77]. The regularity of daily activities, is related to sleep quality, and the lifestyle regularity of the patients with Parkinson's disease is impaired [78].

Consequently, to some extent, the lifestyle regularity can be seen as an approach to indicate the status of human health and the well-being throughout one's life. In addition, Monk et al. proposed a conceptual model trying to explain the interaction between regular lifestyle and health status (Figure 2.9). This model suggests that highly regular behavioral rhythms are probably associated with various of effects, thus leading to restful nocturnal sleep and productive performance of daytime. At first, regular lifestyle leads to regular routine of the day, which is able to affect the photic zeitgebers, so as to the entrained circadian rhythms. As a result, health sleep, mood, and daytime functioning can be enhanced. Furthermore, it can be derive that healthy sleep, mood, and daytime functioning can enhance the regular photic zeitgebers and regular sleep/wake cycle in reverse, thus enhancing well entrained circadian rhythms. From the description it appears that the extrinsic factors and the intrinsic factors form a circle and can interact with each other.

2.2.3. Health-related Quality of Life

To illustrate the concept of quality of life, Torrance used a model to depict the lifetime of an individual (see Figure 2.10), where the lifetime was considered as consisting of two components, quantity of life and quality of life. In the model, a value of quality of life can be measured at any point of quantity of life. Obviously, extending both life span and quality of life can contribute to maximum an individual's living. Many factors may impact on the quality of life, and in particular, health is one of the most important. The concept of health-related quality of life (HRQOL) encompasses those

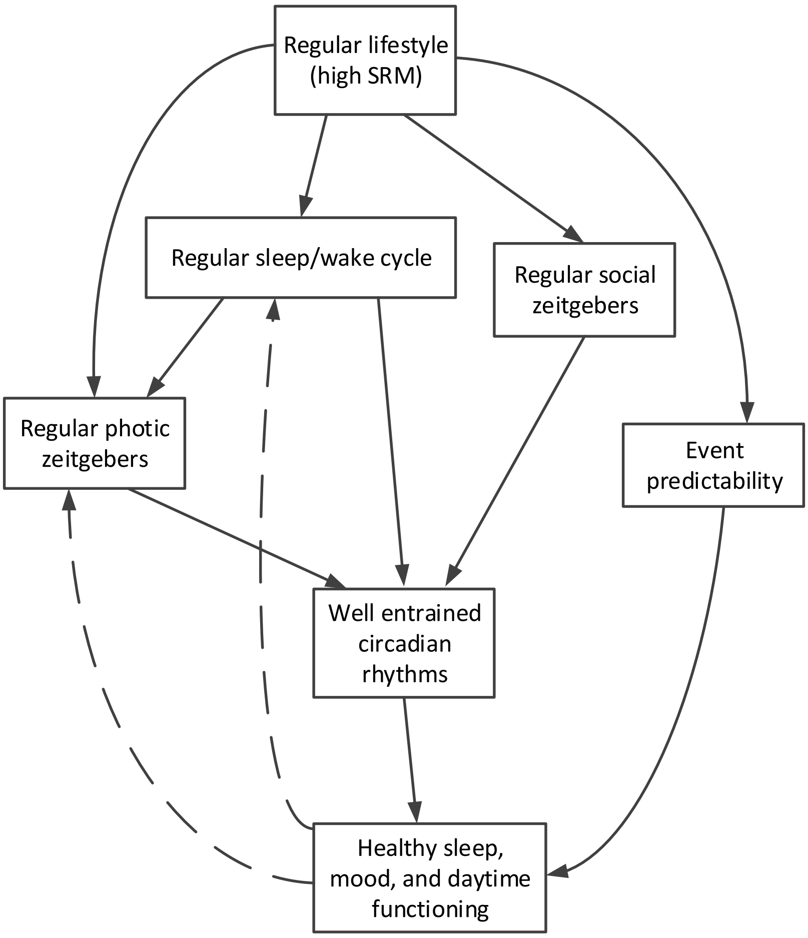


Figure 2.9.: **Conceptual model of how a regular lifestyle might be beneficial to health and well-being (modified according to [74]).**

aspects of overall quality of life that can be clearly shown to affect health – either physical or mental [79].

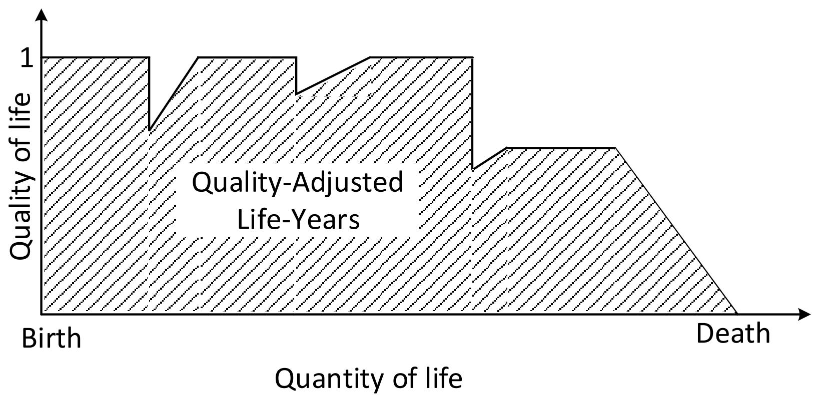


Figure 2.10.: **Quality and quantity of life [80].**

In healthcare, since the acceptance and its measurable outcome, HRQOL has been often used as an indicator to assess patient's health status, and service or interventions in clinical studies. Jones et al. used the outcomes of HRQOL to quantify the magnitude of changes of patients after total hip and knee arthroplasties [81]. The rehabilitation of adults with muscular dystrophy were assessed with HRQOL and ADLs [82]. Juenger et al. assessed health-related quality of life of the patients with congestive heart failure [83]. Mental disorder is also a typical application for this assessment, such as Alzheimer's Disease and dementia [84].

Similar to ADLs and lifestyle regularity, quantification is also necessary for the measurement of HRQOL. Currently, several indexes are often-used, such as SF-36 health survey, EQ5D, HUI (health utility index), and BRFSS [85, 86, 87]. Comparing with the assessments of ADLs and lifestyle regularity, they are more comprehensive. Not only are the items corresponding physical performance investigated, but also the items referring to mental evaluation, emotional feeling, and pain assessment, and sometimes social activities are included as well.

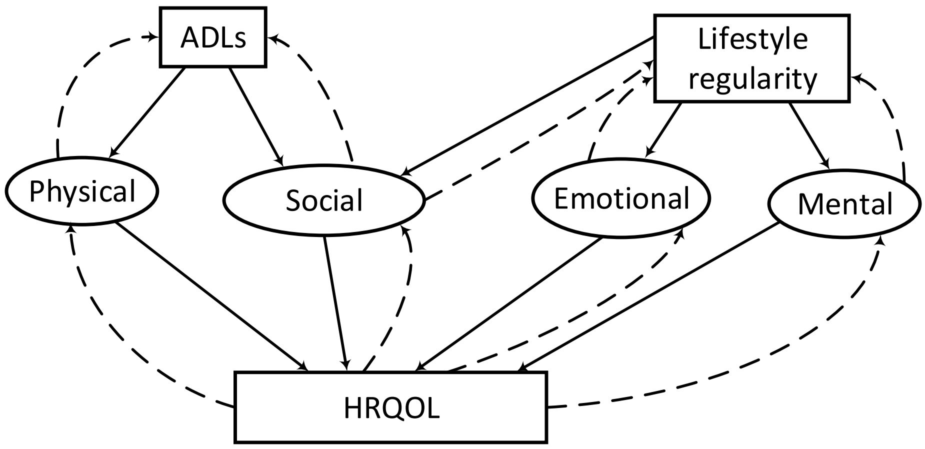


Figure 2.11.: **The interaction between ADLs, HRQOL and lifestyle regularity.**

From the above introduction it can be derived that the change in human health status is reflected in many aspects and they also interact with each other. Both ADLs and HRQOL cover the physical performance. While the ADLs and HRQOL focus more on what the individual is able to do, the lifestyle regularity focuses more on how the individual does some activities. Here a graph is used to illustrate the connections of these three concepts Figure 2.11. On the one hand, the changes in activities of daily living imply the changes in either physical or social activities, which are also the main indicators of health-related quality of life. The change in lifestyle regularity results in the variance of circadian rhythms, thus evidently influencing the emotional and mental modes which are also the main indicators to health-related quality of life. On the other hand, there are also reverse influence. The change in health-related quality of life is able to affect the other two through its components, either positively or negatively.

2.3. Mathematical Background

According to section 2.1, the dataset measured through in-home healthcare monitoring is in the form of temporal sequence. The temporal data mining methods are necessary if useful information is expected to be extracted. In this section, some related mathematical theories are introduced, which mainly refer to temporal data mining.

2.3.1. Association Rules

The term *association rules* (AR), which was firstly introduced in [88], is an implication expression $X \Rightarrow Y$, where X and Y are sets of items. Because the prototypical application of AR is in analysis of sales data, some terms related to marketing are often used to describe it, such as *basket* data and *transaction*. Intuitively, $X \Rightarrow Y$ means the transactions of the database which contain X tend to contain Y. AR has a rang of application domain, from decision support to telecommunication alarm diagnosis and prediction [89]. However, since the dataset of living environment can also be interpreted in the similar way, so far some researches on human activity analysis have been carried out based on AR [90, 91, 92].

Following is the formal statements of the problem: Let $\mathcal{I} = i_1, i_2, ..., i_m$ be a set of items, and $\mathcal{D}$ be a set of transactions, where each transaction T is a itemset such that $T \subseteq \mathcal{I}$. A set of items $X \subset \mathcal{I}$ is called an *itemset*. A transaction T *contains* an itemset X, if $X \subseteq T$. To evaluate an association rules $X \Rightarrow Y$ in a transaction dataset $\mathcal{D}$, three parameters are main indexes, i.e., *support*, *confidence*, and *lift*. More precisely, based on the definitions, the *confidence* denotes the proportion of itemsets containing X and Y of the itemsets containing X; the *support* denotes the proportion of the itemsets containing X and Y in all dataset $\mathcal{D}$; the *lift* denotes the correlation between X and Y. In addition, if *lift* > 1, then it means left side, L and right side, R are relational individuals. The bigger the $lift$ value is, the more correlation between L and R. And their formal definitions are given as following [88, 89].

Confidence c :

$$\frac{num(containing(X \cup Y))}{num(containing(X))} = c\% \tag{2.1}$$

Support s :

$$\frac{num(containing(X \cup Y))}{num(D)} = s\% \tag{2.2}$$

Lift l :

$$\begin{aligned} l(X \Rightarrow Y) = l(Y \Rightarrow X) \\ = \frac{confidence(X \Rightarrow Y)}{support(Y)} \\ = \frac{confidence(Y \Rightarrow X)}{support(X)} \end{aligned} \tag{2.3}$$

2.3.2. Apriori Algorithm

For the problem stated above, discovering all association rules with sufficient support and confidence can be decomposed into two sub-problems [88]:

- Find *large itemsets*, which are combinations of items that have transaction *support* above minimum support.
- Using the large itemsets, generate desired rules with *confidence* above minimum confidence.

As for the first sub-problem, one basic and classic approach is *apriori algorithm*. The algorithm is a multiple passes over the data [89]. In the first pass, the support of individual items are counted and determined which of them are large, hence this is a breadth-first search. Each pass is started with a *seed* set of large itemsets, and the seed set is used for generating new potentially large itemsets, called *candidate* itemsets. At the end of the pass, the actually large candidate itemsets are determined, and they become the seed of next pass. This process continues until no new large itemsets are found.

In apriori, the basic combinatorial property is adopted, which is any subset of a large itemset must be large. Therefore, the candidate itemsets having k items can be generated by joining large itemsets with $k-1$ items, and deleting those that contain any subset that is not large, thus leading to generation of a much smaller number of candidate itemsets, i.e., in effect pruning the search space. For a formally statement, the pseudocode description is given (see Algorithm 1) [89]. As to the second sub-problems – generating rules, for a large itemset l, suppose a is a subset. For every a, output a rule of the form $a \Rightarrow (l-a)$, if the ratio of support(l) to support(a) is above *miniconf*. Particularly, towards implementation, Borgelt et al. carried out apriori algorithm based on *pre-fix tree* concept [93]. Each node n_S denotes a counter for an itemset S. The counters for the itemsets are denoted in a special kind of *pre-fix tree*, whereby counters can be stored efficiently and In the addition, *pre-fix tree* supports processing the transactions and generating the rules. During the first step of apriori algorithm, *pre-fix tree* is created level by level, and some branches of the tree can be pruned according to *minimum support* as well.

For applying the methods, one can easily inputs *transaction* dataset and some parameters restricting the discovering process, including the *minimum length*, *targets* and *minimum support* or *minimum confidence*. And the outputs are the association rules and their *lift*, *support* and *confidence* w.r.t. *target* = "rules", or frequent itemsets and their support w.r.t. *target* = "frequent items".

Algorithm 1: **Apriori algorithm**

$L_1 = large1 - itemsets$;
for $(k = 2; L_{k-1} \neq \phi; k++)$ **do**
    $C_k = apriori - gen(L_{k-1})$; ▷ new Candidates
    for all transactions $t \in \mathcal{D}$ **do**
        $C_k = subset(C_k, t)$; ▷ Candidates contained in t
        for all $candidatesc \in C_t$ **do**
            $c.count++$;
        end for
    end for
    $L_k = \{c \in C_k | c.count \geq minsup\}$
end for
$Answer = \bigcup_k L_k$;

Notation:
k-itemset: An itemset having k items
L_k: A set of large k-itemsets (those with minimum support). Each member of this set has two fields: i): itemset and ii): support count
C_k: A set of candidate k-itemsets (potentially large itemsets). Each member of this set has two fields: i): itemset and ii): support count.
$\hat{C}_k$: A set of candidate k-itemsets when the TIDs of the generating transactions are kept associated with the candidates.

2.4. GAL-NATARS Study[1]

2.4.1. Motivation of the Study

The older patients who suffered from hip fracture have a high percentage of complications, such as osteoporosis and unbalanced gait. Due to the limitation to the mobility, a walker or a cane is usually necessary. In addition, they are at risk of future deterioration. It is also necessary to provide them the intensive and individually tailored observation, especially for those living alone. Earlier discovery of adverse trend of health status is expected, by which the earlier intervention can be implemented. In the context of real-life situations, the conventional healthcare after discharge from hospital is conducted without the support of technical assistant systems, where only snapshots can be provided. The sensor-enhanced living environment has the potential to provide continuous monitoring of patients' behavior during their daily living.

Therefore, the GAL-NATARS study (which stands for "Nutzung von assistierenden Technologien für das Assessment von Risikoprofilen bei Patienten mit stattgehabter Schenkelhalsfraktur") focuses on the group of patients who suffered from hip fracture. As the target population, the observing results from those patients are thought to be representative. The primary goal of this study is to investigate the technical feasibility of the usage of health enabling technologies to support independent living at home and the acceptance by users, including the patients and the caregivers.

The data obtained makes it possible to detect an occurrence of deterioration of the function and the self-care of the patient. It may also succeed in discovering newly emerging risk situations. Believing that their daily behavior is the indicator of rehabilitation and health status after discharge, some relations between their behavior and health status are expected to be revealed. With respect to this thesis, the data acquired in this study is thought to be appropriate to describe the patients' behavior patterns and profile the patients' lifestyle, thereby assess the patients' health status and rehabilitation progress. Therefore, the datasets of this study are chosen as the data source to validate the methods proposed in the current work.

[1] This section is based on the study plan of GAL NATARS [94], and only the related content is introduced.

2.4.2. System Design Criteria

2.4.2.1. Study Type

The GAL-NATSRS is a study in terms of technical feasibility and acceptance of users and families, and the usability for the assessment of risk profiles in patients will be studied. It is a multi-center study. The recruited subjects live in three cities of Lower Saxony, including Braunschweig, Oldenburg, and Lingen, and the monitoring processing is conducted in their real living environments, either apartment or house. The data is processed and stored in Hannover Medical School.

2.4.2.2. Study Duration

The total duration of the study is 12 months, and the observation period of each subject is 3 months.

2.4.2.3. Inclusion Criteria

In this study, subjects included should satisfy the following criteria:

- Being treated with osteology synthetic (osteosynthetic) or endoprosthesis, and participating in a geriatric rehabilitation after hip fracture;
- Full weight bearing;
- Age ≥ 70;
- Living within 60 km radius of one of the geriatric hospital;
- No pet with the resident;
- Living alone (including assisted living);
- Mini-Mental-State Test ≥ 20.

2.4.2.4. Subjects Recruitment

According to the inclusion criteria, the subjects were recruited from the patients who have been treated in a hospital geriatric department for acute or rehabilitative completion of femoral fracture with osteosynthesis or joint replacement. Following the hospitalization with or without nursing support, they move to a independent home, i.e., own apartment with no special assistance features, but not a nursing home with assisted living.

The recruitment of the volunteers are the patients of the following three institutions: 1). St. Bonifatius Hospital Lingen; 2). Geriatrisches Zentrum Oldenburg; 3). Städtisches Klinikum Braunschweig, Klinik IV (Geriatrie).

At the end, a sum of 30 subjects are included, each of whom is observed for 3 months. Because of the exploratory nature of the observational study, no sample size estimation is conducted.

2.4.3. Study Procedure

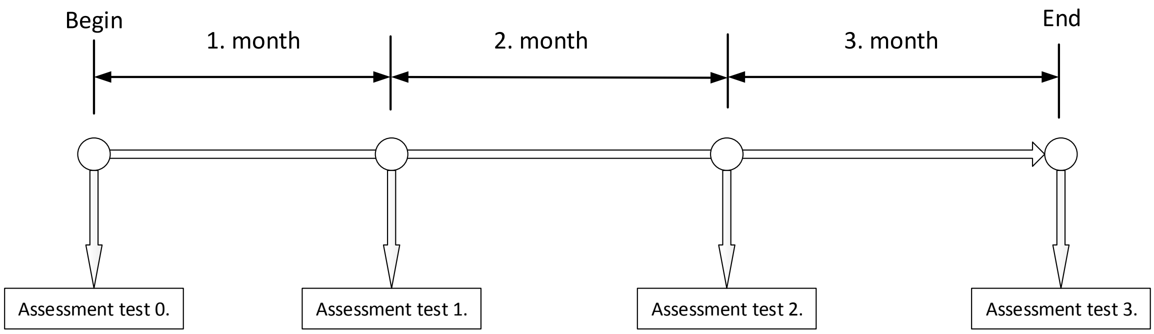


Figure 2.12.: GAL-NATARS study procedure.

First of all, after the subjects are discharged from hospital, complete geriatric assessments are conducted on each of them. Those assessments are repeated at the end of the 3-month observation. In addition, a selected set of assessments are also performed every month (see Table 2.1). The results of the assessments are used to be compared with the data collected through the sensors of the assistant system (i.e., health enabling technologies).

The assessments include[95]:

- Activities of Daily Living (ADL), Barthel Index
- Instrumental activities of daily living (IADL) according to Lawton and Brody

Table 2.1.: Monitoring Plan

Measuring Parameter	time point t_0	time point t_{M1}	time point t_{M2}	time point t_{M3}
Barthel-Index	+	+	+	+
IADL	+	+	+	+
VAS	+	+	+	+
MMSE	+			+
Time-set test	+			+
GDS	+			+
Body weight	+	+	+	+
BMI	+	+	+	+
MNA	+			+
Hand strength	+	+	+	+
Tinetti I	+	+	+	+
Tinetti II	+	+	+	+
TUG	+	+	+	+
SPPB	+	+	+	+
SA	+			+
ETB	+			
INT	+			+

Note:
IADL: Instrumental Activities of Daily Living;
VAS: Pain Assessment Visual analogue scale;
MMSE: Mini Mental State-Examination based on Folstein;
GDS: Geriatric Depression Score;
MNA: Mini Nutritional Assessment;
TUG: Timed up-and-go-Test;
SPPB: Short Physical Performance Battery;
SA: Social assessment;
ETB: Scale for the Assessment of Technology Readiness (in German: Skala zur Erfassung der Technikbereitschaft);
INT: Personal Interview.

- Mini-Mental State Examination based on Folstein with Time test
- Tinetti test with the components I and II
- Timed Up-and-Go test
- Hand strength
- Body Mass Index (BMI)
- Mini Nutritional Assessment(MNA)
- impedance measurement
- Pain Assessment Visual analogue scale (VAS)
- Social assessment
- Technology acceptance [96], usability [97]

2.4.4. Sensor Operation Mechanisms

Basically, a sensor is a converter that measures a physical quantity and converts it into a signal that can be interpreted by observer or electronic instrument. It is the way for device to sense the environment, and it is also an approach to extend human sensing domain. As discussed in chapter 1, for practical implementation of in-home healthcare observing, only the unobtrusive sensors are adopted. Here the unobtrusive sensors are such sensors that can only obtain limited information and are deployed in specific positions with no obvious influence on the subject's daily life as usual.

Some kinds of sensors are used in GAL-NATARS study for human behavior related detection. Based on specific operation mechanisms, they are deployed in some specific positions. In order to explain their positioning plan, and for the purpose of understanding the data format, it is necessary to introduce the sensing mechanisms of the sensors used in the study. In this section, only a few sensors, the data of which is used in this thesis, are covered, including presence sensor, vibration sensor, and contact sensor.

2.4.4.1. Presence Sensor

From the name, it can be intuitively implied that the presence sensor is used to detect the presence of a object arising in the sensing area. Passive Infrared (PIR) Sensor (Motion Sensor) is a typical electronic sensor that measures the presence of a object through measuring the change of infrared (IR) light inside the sensing area [98]. The sensor is small, inexpensive, and low-power and commonly used in homes and business. Because human body usually generates more amount of infrared than the environment, this sensor is suitable to detect human behavior at home.

The sensing component consists of a lens and a set of sensors. There are two slots with the sensor set. The slot is made from the pyroelectric materials that is sensitive to IR, which can detect levels of infrared radiation and generate energy when exposed to heat. When the sensor is idle, both slots detect the same amount of IR, i.e., the ambient amount radiated from the room or walls. If a person passes by, the first slot is generates a pulse, referring to *positive differential*, and when the person leaves the sensing area, a *negative differential* is generated. These changing pulses are the signal that can be detected (see Figure 2.13).

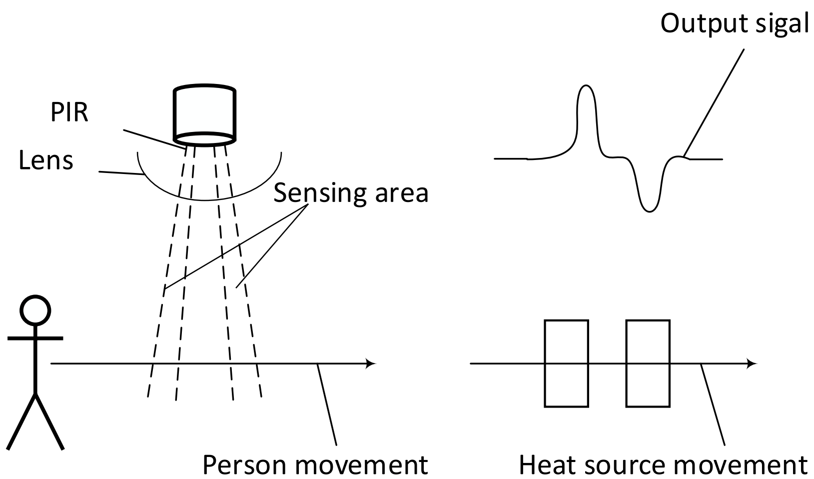


Figure 2.13.: The sensing mechanism of PIR sensor [98].

Towards extending the sensing area, a Fresnel lens is usually adopted to cover the sensor. Taking the PIR sensor module (FS20-PIRA) used in GAL-NATARS study as an example, the sensing area is demonstrated in Figure 2.15, where the coverage angle is 90° and the sensing range is about $6meters$. Furthermore, the lens is split into multiple sections for a scattering of multiple small areas (see Figure 2.14). That means when a person moves from one small area to another, the sensor will be triggered.

Considering the sensor module FS20-PIRA, its technical data is listed in Table 2.2. The FS20-PIRA module also contains a control component, including an supporting

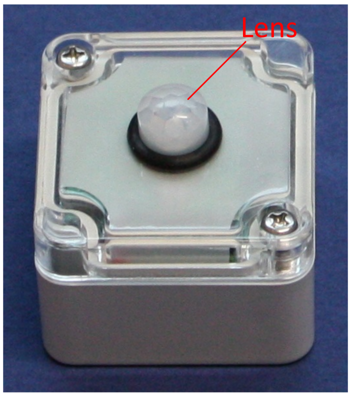


Figure 2.14.: The Lens of FS20 PIRA.

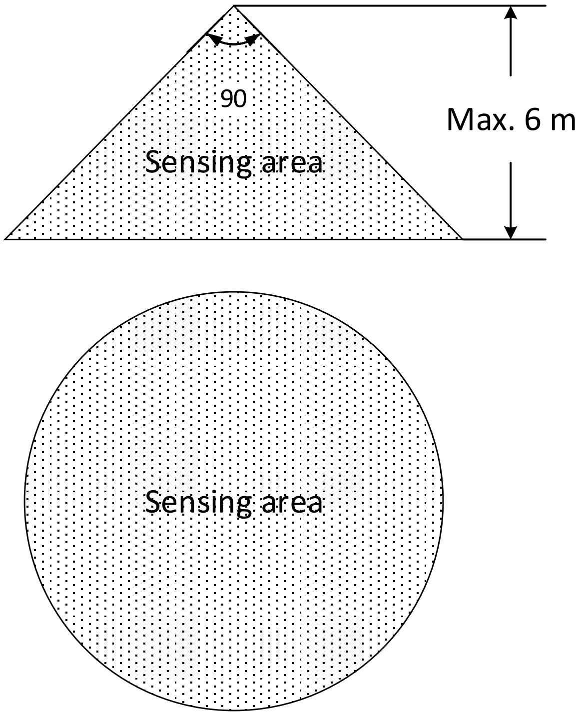


Figure 2.15.: The sensing scope of passive infrared sensor. This figure is modified from the manual of motion sensor (Type: FS20 PIRA) [99]

and control circuit and a micro-controller, which can interpret the signal and transfer it to a base station for further processing. For energy saving, a minimal transmitting period T_0 is set, which means the minimal time distance between two samples is T_0. If a person performs any motion in the sensing area at t_1, and finishes the motion at t_2, and $t_2 - t_1 < T_0$, then the circuit would not detect any event.

Table 2.2.: The technical data of FS20-PIRA [99]

Function principle	passive infrared detection
Coverage angle	about 90°
Number of channel	2
Range	about 6m
Frequency	868.35 MHz
Radio range	radio range: 100m (free field)

2.4.4.2. Contact Sensor

For detecting some manipulations on objects, contact sensor is frequently chosen. Reed switch sensor, is an electrical switch, which is mostly used to detect the change of state. A reed switch consists of a pair of contacts that are contained in a glass hermetic shell (see Figure 2.16). One end of the contacts is fixed, while the other end is covered with some electro-conductive material and can move freely under the effect of an external magnetic field [100]. Once there is magnetic field applied to the switch, the free end moves to the fixed one and the switch is turned on, so that a signal can be detected.

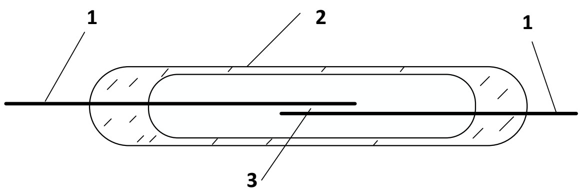


Figure 2.16.: Construction of a modern reed switch. 1 – contact elements (springs) from permalloy; 2 – glass hermetic shell [100]; 3 – working gap. Once there is magnetic field applied on the switch, the working gap disappears and there is current passing.

In GAL-NATARS study, the sensor module FS20-TFK, which contains reed switches inside, was used to detect the interactions with some doors. Whenever the observed door is opened or closed, FS20-TFK will make a record of the event and transfer it to the base station in this study.

2.4.4.3. Vibration Sensor

Vibration sensor is a device that can detect the vibration of an object. It can be implemented in a variety of ways, e.g., electronic accelerometer and mechanical sensor. It is usually used for glass break detector, laptop protection, or anti-theft device. For the motion detection at home, it can be mounted on some objects that can generate obvious shake if they are manipulated by user. In GAL-NATARS study, a kind of mechanical sensor vibration sensor (FS20 ES1) is used to detect sleeping (mounted at the bed), manipulation with wall-cupboard and drawer in the kitchen.

The basic mechanical structure of this type of sensor is demonstrated in Figure 2.17. As shown, the sensing component consists of a metal ball and two contact plates. The mental ball is embedded between these two contact plates. As long as the ball is in the middle of the hole, there is no connection between the two contacts. Once there is shake, the ball connects the contacts and the sensor is triggered [101].

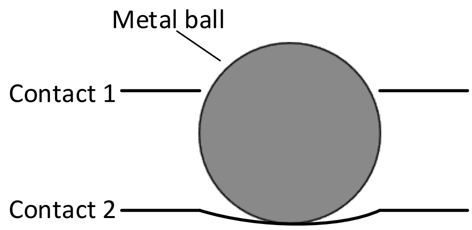


Figure 2.17.: The sensing mechanism of vibration sensor based on metal ball [101]. As a connector, the mental ball is embedded between two contact plates.

The alarming transmitting interval T_s can be adjusted, which means once T_s is set, the devices ignores any shake within T_s after the last triggering. Therefore, this sensor can be used to extract both interval-based events and time point-based event, which are defined in section 3.2. For the detection of sleeping, if there are triggers of the sensor on the bed in certain time, implying that a person is in the bed, and then the starting time and ending time can be inferred. As to the manipulations with the drawer and cupboard, each trigger can be seen as one manipulation.

2.4.5. Measure Devices Installation

The monitoring system is based on the application of sensing technology and wireless communication technology. To make use of these technical components in the geriatric context, firstly a suitable infrastructure in the living environment of the subjects needs to be built.

For the acquisition of relevant parameters of geriatric patients with hip fracture in the home environment, the sensors introduced above are installed in each apartment in specific positions which are unobtrusive for human daily living. A sum of 5 kinds of sensors are used in this study, including motion sensor, switch contact, vibration sensor, current sensor, and accelerometer. Motion sensor is used for detecting the subject's presence in sensing area, for instance, the presence in a specific room or in a one specific area in a room. Switch contact and vibration sensor are used for detecting interactions with objectives, which are mainly the usage of doors, windows, and fridge etc. Their functions are summarised in table 2.3. All the sensors constitute

Table 2.3.: Sensors and Devices that are used

Name	Description	Hints
FS20 TFK door and window contact	simply contact switch with wireless connection, battery operated	connected to FHZ1000PC
FS20 ES1 shock sensor (vibration sensor)	record vibrations, battery operated	connected to FHZ1000PC
FS20 PIRA motion sensor	Infrared motion detector, battery operated	connected to FHZ1000PC
FHZ1000PC	component for receiving the radio data	connected with the computer through USB
GAL computing system (arithmetic system)	Minicomputer with GAL-Software-Components for the connection of sensors and storage of the data	
"Plugwise" current sensor (smart meter)	decentralized power sensors (smart meters)	inserted between the equipment and power outlet

a near body area network through wireless communication, which is FS20 bus system in this study. Through the sensors installed in each room in the living environment, each single event with time stamp is recorded, e.g., a passing through a sensing area of a motion sensor or one time operation on the fridge door. For further processing, the acquired data is transmitted towards a base station with a FHZ1000 PC receiver, which is in charge of data collecting, formatting, and storage.

Figure 2.18.: Example of an apartment layout.

A common apartment includes some functional components, such as living room, sleeping room, toilet, and kitchen. In addition, some appliances are located in different zones. In order to demonstrate the deployment of sensors for in-home monitoring, an example layout is given (see Figure 2.18). This figure shows that all the rooms are detected by the motion sensors (green boxes) mounted on the door frame, and the whole apartment can be seen as five zones, including four blue shadows and the bedroom without motion sensor; two doors (main entrance and door to the balcony) and fridge are detected by the switch contacts (yellow boxes); Bed and the drawer and cabinet in the kitchen are detected through vibration sensors (red boxes). Besides, some appliances, such as television, microwave, lamps, heater and so on are detected by electric current sensors. For subjects lacking of mobility, most of their activities are performed within their apartments that has limited area and most of their behaviors are interactions with those objectives within it.

2.5. Chapter Summary

This chapter mainly introduced the basic background of knowledge and method.

First of all this chapter reviewed the in-home health monitoring, including its definition, structure, application and users, and in addition the data processing methods frequently used for in-home health monitoring – intelligent data processing (IDA) – was also covered. Considering the interdisciplinary character, the medical background and mathematical theories are described. Finally, the GAL-NATARS study, where the analysis material data comes from, is systematically introduced.

Chapter 3

Abstractions of In-home Scenarios for Behavior Analysis

Living environments play an important role in our daily life. Apart from working activities, mostly we perform the others inside our apartments or houses. For those frail elder, they spend even more time at home. In the current work the term *in-home scenarios* is adopted to describe the situations and conditions related to living environments. When considering in-home scenarios, two factors have to be taken into account, i.e., the living environment and the resident. The living environment provides the basic conditions, and the resident carries out behavior under these conditions. In other words, the interactions between the environment and the resident constitute the *in-home scenario*. In real-life situations, diverse activities occur w.r.t. individuals. Abstracting the substantive characteristics of these situations is helpful for extracting useful information.

Towards formally describing the resident's behavior, this chapter presents the abstractions of in-home scenarios from these two basic factors. Firstly the living environment is mapped onto a graph-based model so as to represent the topology and structure; secondly, the basic elements of activities are abstracted as *events*. At last, human behavior patterns in the form of event distribution over time-line are abstracted by means of a hierarchical format.

3.1. Modeling Living Environments

Despite a variety of living environments with various architectures, some common characters can still be identified. Firstly, two basic elements constitute the living environment, including *zones* and *objects*. Secondly, the similar function units are mostly involved in different styles of living environments. Here *zones* refers to relatively large areas, such as rooms and certain areas, and for fully facilitating daily living, a range of *objects*, which are usually the furniture and the electric appliances, can be located in zones. Between zones are *connectors*, referring to doors and passages, whereby the resident can pass through from one zone to another. For residents living in it, one can reach any zone within the living environment from a certain point. Moreover, no matter how the architecture of a living environment is, the essential function units are always included, such as living room, bedroom, bathroom, and kitchen etc. It has to be pointed out that these function units are designed in different styles, with various areas and equipments, thus resulting in different topologies.

Based on those common characteristics, in this section a graph-based model, which is named as *@home graph*, is proposed to formally describe the living environment. The basic idea of *@home graph* is to map the elements in living environment scenario onto the nodes, and to map the relationships between elements onto edges. Two types of nodes are defined in *@home graph*, i.e., *backbone node* and *leaf node* (see Def. 1 and Def. 2). Note that the terms "big area" and "small area" in the definitions can not be precisely defined since they closely depend on the application background.

Definition 1 ***Backbone node:*** *Backbone node is defined as the node that denotes zone, which is relatively big area, such as a room and specific area.*

Definition 2 ***Leaf node:*** *Leaf node is defined as the node that denotes object or position, which is relatively small area or point, such as the usage or operation on a object, or a specific position of a big area.*

Figure 3.1.: **The example apartment layout(The material is from the GAL-NATARS study).**

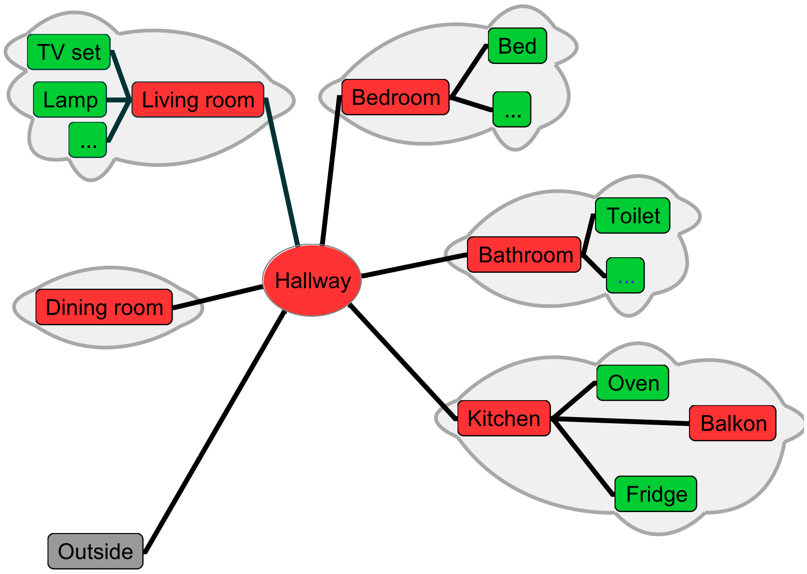


Figure 3.2.: The @home graph based on example apartment layout.

In order to explain the generation of *@home graph*, we reuse the example apartment layout (see Figure 3.1). This is a normal apartment, and there are a number of zones in it corresponding to the function units, including living room, bedroom, dining room, kitchen, bathroom etc., which are denoted by brown shadows. According to the definition of *backbone node*, in this case these zones are mapped onto the backbone nodes (red bubbles in Figure 3.2). There are doors as connectors, which are denoted by the edges between backbone nodes. The objects placed in zones are mapped onto the leaf nodes (green bubbles in Figure 3.2). The relationships between *backbone node* and *leaf node* are decided by the location of corresponding objects. If one object is inside a zone, the related leaf node belongs to the backbone node, and the relation is also denoted by the edge, and the *leaf node* is also called the *child node* of the *backbone node*. Besides, one *backbone node* and its *child nodes* can be further *grouped*. Through *@home graph*, it is obvious that the architecture is represented by backbone nodes of *@home graph*, and the *leaf nodes* complement and enrich the representation.

Once the graph has been generated, its properties have to be clarified. There is no *leaf node* belonging to two or more *backbone nodes* at the same time. But it is possible for a *leaf node* to transit across different groups, for instance, when the object is moved from one zone to another. For two backbone nodes, if they are connected directly by one edge, say they are *neighbours* with respect to each other. Besides, each element in the graph has its *state*, which can be *active* or *inactive*.

Generally, the *@home graph* can briefly represent the information of living environments. In the context of pervasive computing, various parameters can be measured by setting appropriate sensors at appropriate positions, thus building a sensor-enhanced living environment. Given the *@home graph*, regarding to specific application, a set of components, either nodes or edges, can be chosen to be measured through available sensors. In the current work , the form $C_i \vDash S_j$ is adopted to denote that the component C_i is detected by the sensor S_j. Theorem 1 decides the *state* of one element. Apparently, an active component indicates occurrence of the objective activity or event.

Theorem 1 *if (S_i is triggered at time t) $\wedge$ ($C_j \vDash S_i$) , then the state of component, state(C_j) $\equiv$ active.*

Consider the following case. Suppose there is one resident living in the example apartment and some sensors are deployed to detect all backbone nodes and all the leaf nodes mapped from electric appliances. If he or she moves from the living room to the kitchen and cooks for a while and then moves out of the kitchen, then the *active* state on *@home graph* should transit from the group of living room, through hallway group, to the kitchen group, and then turn back to the hallway group. From this case, we can easily find that the shift of active *state* over the elements in *@home graph* can reflect the resident's change of position, i.e., trajectory.

In summary, the living environment can be formally described by a *@home graph*, where special components are defined w.r.t. various elements of living environments. Not only are the static topology and structure represented, but also dynamic occurrence of interaction between residents and environments are manifested by *state* change of components. What makes the abstraction more significant is that it can contribute to information extraction that will be introduced in the following chapter.

3.2. Abstractions of Events

If a certain activity is wished to be accomplished, three essential factors are needed, including some actions, a certain position (e.g., specific room), and the time duration (i.e., time interval).

In view of the practical context of in-home healthcare observing, some temporal data mining methods are often used to discover patterns for lifestyle description. Thus, for data processing and information extraction, an uniform description is necessary. In the previous section the living environment has been modeled by the *@home graph* and human behaviors are reflected by the *state* change of components in *@home graph*. Note that in the *@home graph*, what are detected are the changes of components, rather than human activities. The direct information of the living environment is not exactly what activities appear when there are *state* changing. However, the change of *state* means that some kind of related activities are emerging. It is feasible to indirectly assess human behavior though the change of state. In this section the term *event* is used to abstract the *state* change of components. In the *@home graph*, a sequence of *events* are collected after a certain time interval measuring of a living environment. The activities over the whole measuring duration have influence on the record of events in time order. It has to be pointed out that the abstractions based on the sequence of events are different from the abstractions of normal time series. The main task of abstraction of time series is to labeling the time series such as the trend of the numeric value and the intervals with the approximative value, while the abstractions of events means to abstract the describing manner of events.

The *event* consists of two attributes, i.e., event identifier (ID) and occurring time, denoting which component is active and when it is active, respectively. For unifying the form, the event, say E, is denoted as $[e : t]$, where e is event ID, indicating which kind of event it is, and t is time attribute, indicating when the event occur. From the point of view of time attribute, some activities are accomplished instantaneously, while some others need much more time. As an example, the action of opening the window can be finished within one second, while using toilet probably persists several minutes. In the former case, the time duration spent on opening the window can be ignored in that there is nearly no effect on understanding this event. In the latter case, however, as for the toilet usage, the duration has to be taken into account as it may reflect some health status of the inhabitant. Based on the difference reflected on the time property, events are grouped into two categories, i.e., *time-point-based event* and *time-interval-based event*.

Time-Point-based Event

Any event has its starting time and ending time, say $t.s$ and $t.e$, respectively. Intuitively, the so called time-point-based event is such that occurs at a time point and its duration is not important for information extraction and can be ignored. Time-point-based event is subject to two constraints:

- it is accomplished in a relatively short duration, and more formally, $t.e - t.s < \varepsilon$, where ε is a extremely small threshold defined based on application.
- there is no other event occurring between $t.s$ and $t.e$.

More formally, time-point-based events can be denoted in the form $E[e : t]$, and $t :=< t.o >$ is event occurring time, which can be any value between $t.s$ and $t.e$, subjecting to $t.s \leqslant t.o \leqslant t.e$.

Time-Interval-based Event

The time-interval-based event is such that occurs over certain time duration that can reflect variance of the event and cannot be ignored. Similarly, the time-interval-based event has to satisfy the following constraints:

- it is accomplished in a relatively long duration, i.e., $t.e - t.s > \varepsilon$, where ε is the threshold defined based on application.
- there may be other events occurring between $t.s$ and $t.e$, either time-point-based or time-interval-based.

More formally, time-interval-based event can be denoted in the form $E[e : t]$, and $t :=< t.s, t.e >$ denotes the starting time point and the ending time point.

People's life usually follows a kind of periodic way with respect to time. For instance, we sleep at night, and we eat a meal at the similar time during a day, and there is rush hour in the morning and in the evening on weekdays, which is not so significant at the weekend. Therefore, the time-line can be regarded as duplicates periodically generated. As for specific background, this may be in the form of daily, weekly, monthly, etc.

Suppose there is an events database $\mathcal{D}$ through a time span $\mathbf{S}$, to analyze the lifestyle of the resident, the database is split into a number of sections based on the period P. The *length* of one period is defined as l_P, which means $\mathbf{S} = n \cdot l_P$, where n is the number of periods within $\mathbf{S}$, and $n \in N$. The i^{th} period P_i corresponds to the interval $[i \cdot l_P, (i+1) \cdot l_P]$, $0 \leqslant i \leqslant n-1$. For instance, if the daily lifestyle is wished to be analyzed, then the l_P is set to be one day (24-h). Similarly, the weekly or monthly periods can also be analyzed as long as l_P is set to be 7-day or one month. The set of events occurring in period P_i is denoted as $D[i]$, so the database $\mathcal{D} = \bigcup_{i \in [1,n]} D[i]$. And for any event E_j in period P_i,

$$D[i] = \{[e_j : t_j] | t_j \in P_i\}$$

Note that for easily analyzing inhabitant's lifestyle, the value of occurring time of event E is assorted to be the offset to the beginning of the period P that e belongs to, whereby the difference between events belonging to different periods can be easily compared.

Based on Allen's interval algebra [102], the possible relationship between these two kinds of events can be depicted as: *before/after/meet*, *during/cover*, *overlap*, and *equal*. Since the constraint of time-point-based events, there is no *during* and *overlap* relationships between two time-point-based events, as illustrated in (see Figure 3.3). Considering the particularity of the abstractions of in-home scenario, the relationships of *before/after* and *meet* are regarded as the same. When dealing with the relationship between interval-based events, they can also be regarded as two pairs

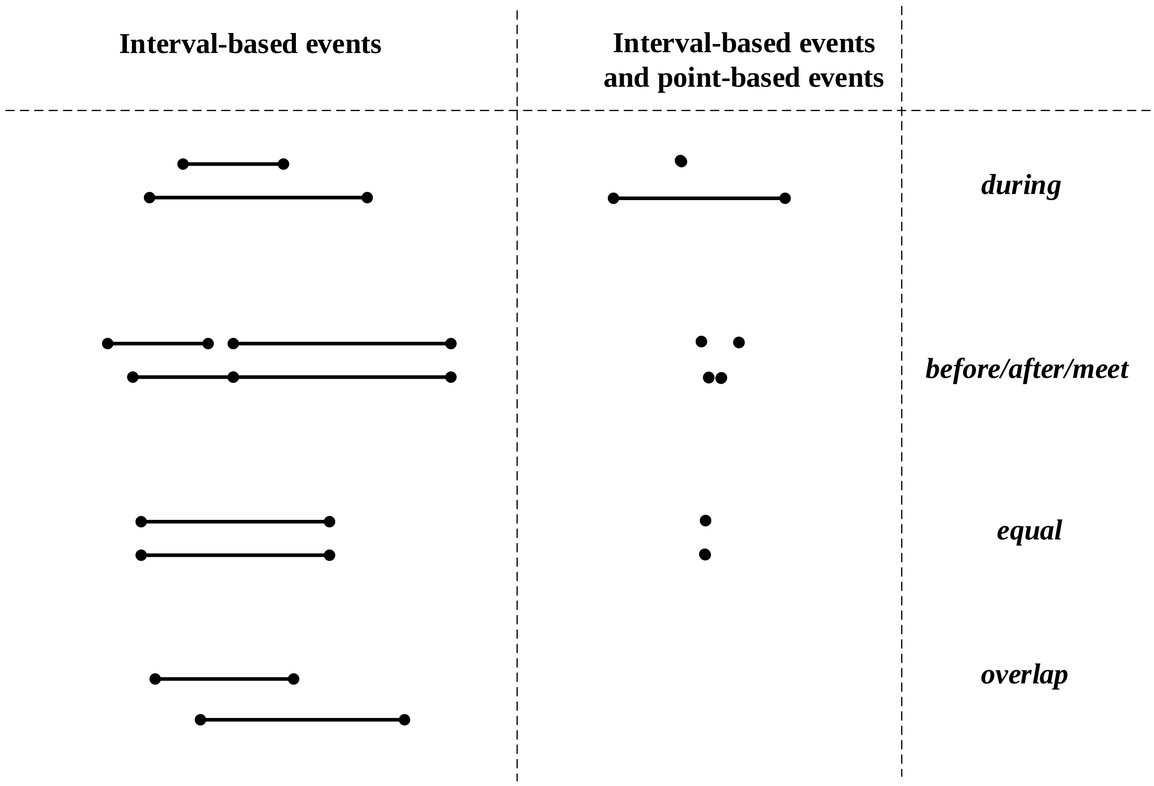


Figure 3.3.: The relationships between events. The line segment denotes the time-interval-based events and the point denotes the time-point-based events.

of point-based events. Take the *equal* of two interval-based events as an example, this can be seen as two *equals* of two pairs of time-point-based events. If there are more than one residents living in, it is also possible to arise *overlap* relationship.

Consequently, based on the concepts introduced above, a database used for discovering lifestyle patterns is constructed. It is evident that the database $\mathcal{D}$ is more structured and can be depicted as Figure 3.4. In each period P_i, the set of events $D[i]$ are ordered according to their time attribute (the offset to the starting of the period the event belongs to), which are occurring time for time-point-based events and starting time for time-interval-based events.

3.3. Abstractions of Patterns

As a result of variance of health status, habits and cultures of individuals, different activities may be performed in same time intervals across periods; and different time intervals and locations may be used to conduct some ADLs as well. Thus say each

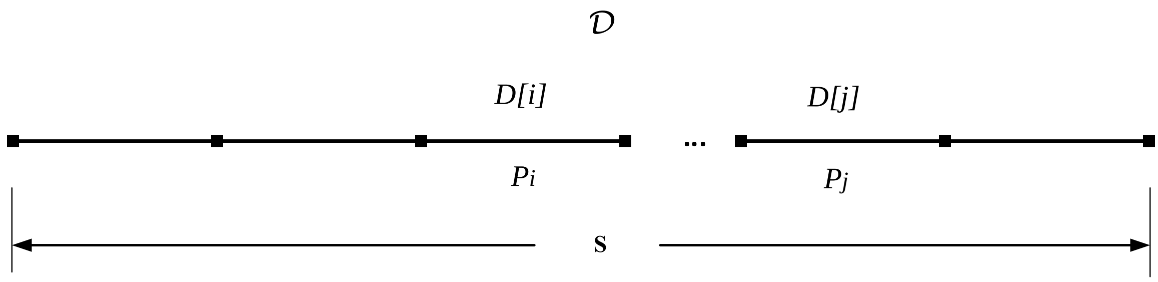


Figure 3.4.: The structure of the database based on the definition of period.

individual has his or her own typical behavior pattern. In this section human behavior patterns will be formally abstracted.

3.3.1. Periodic Events

Over a number of consecutive periods, if a event or a set of events are subject to some time constraint, then say they are *periodic*. As introduced in the previous section, the time attribute of one event is the offset to the beginning of the period. As shown in Figure 3.5, suppose there are two events $E_i[e_i : t_i]$ and $E_j[e_j : t_j]$ where $e_i = e_j$, and E_i belongs to i^{th} period P_i, and E_j belongs to j^{th} period P_j. If the time attributes of E_i and E_j satisfy the following conditions,

$$|t_i - t_j| \leqslant T_{th} \tag{3.1}$$

i.e, for time-point-based events

$$|t.oi - t.o_j| \leqslant T_{th} \tag{3.2}$$

and for time-interval-based events

$$|t.s_i - t.s_j| \leqslant T_{th} \wedge |t.e_i - t.e_j| \leqslant T_{th} \tag{3.3}$$

where T_{th} is the predefined threshold to control the consistency, then say these two events are *periodic*. Considering the time-interval-based events, in other words, if two of them are *periodic*, then both the start and end of these two events are *periodic*.

3.3.2. Hierarchical Abstraction

In the context of pervasive computing, some zones or objects are observed targets. With respect to specific applications, such as health status and culture background of

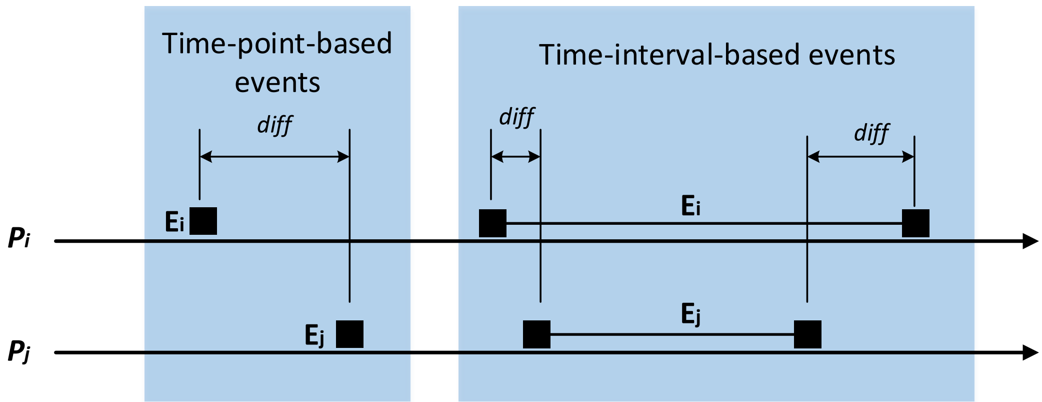


Figure 3.5.: Two pairs of periodic events in period P_i and P_j. The first pair are time-point-based events and the second pair are time-interval-based events. The left-side blue zone illustrates the time-point-based events, and the right-side blue zone illustrates the time-interval-based events.

the resident, the resident may show a range of habits, resulting in different patterns of objects and zones usage. Manifested by *@home graph*, unique event distributions are exhibited w.r.t. individuals. To maximum one specific resident's behavior, a set of events should be selected as indicators. In order to formally describe the human behavior, this section will present a *hierarchical abstraction*.

As already mentioned, people are usually experiencing a periodic lifestyle. Based on this assumption, this thesis will make use of lifestyle regularity as an important indicator to assess the stableness of human health status. This can be manifested by event distribution across periods. First of all, two concepts are defined, i.e., *basic period* (P_b) and *multi-period* (P_{mul}) (see Def. 3 and Def. 4).

Definition 3 *Basic period: the time unit that can present human activity rhythm and the time interval shorter than it is not considered as period.*

Definition 4 *Multi-period: the time interval that is constituted by a number of consecutive basic periods.*

Taking the lifestyle regularity and circadian rhythm into account, in the current work the basic period is set to be *one day* (24-h), and the multi-period is set to be *one week* (7-d). The basic period can be further split into a number of *segments*, some of which may be chosen as interesting time intervals that may contain meaningful information. As shown in Figure 3.6, there are n *basic periods* and each of them is split in the same way, and a number of consecutive *basic periods* make up a *multi-period*. From Figure 3.6 it is obvious that if all the basic periods are put in parallel, the change of event distribution over the same positions of all basic periods is clearly shown through looking into the segments with the same index, which are from consecutive basic periods. For example, the variance in corresponding time interval can be revealed by investigating the distribution over all *segments* of *basic period 1* to *basic period n*. This also explains why the event time attribute is processed to be the offset of the beginning of period. Depending on the application context, it is also possible to set the multi-period as one day, a couple of days, week, month, or even year. By investigating more multi-periods, the longer term trend of human behavior will be obtained. In the current case, because the basic period is set to be one day, and the multi-period is set to be one week, the basic period can be split into intervals such as morning, afternoon, evening, night and late night, so as to present characters of specific intervals. Otherwise, if the basic period is set to be one week, and multi-period is one month, then the week can be split into workday and weekend.

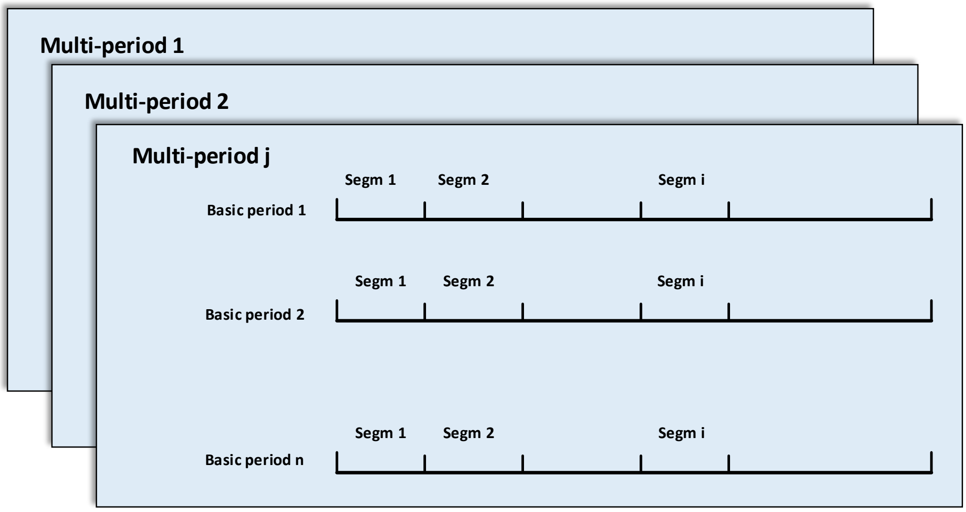


Figure 3.6.: **The hierarchical time-line.**

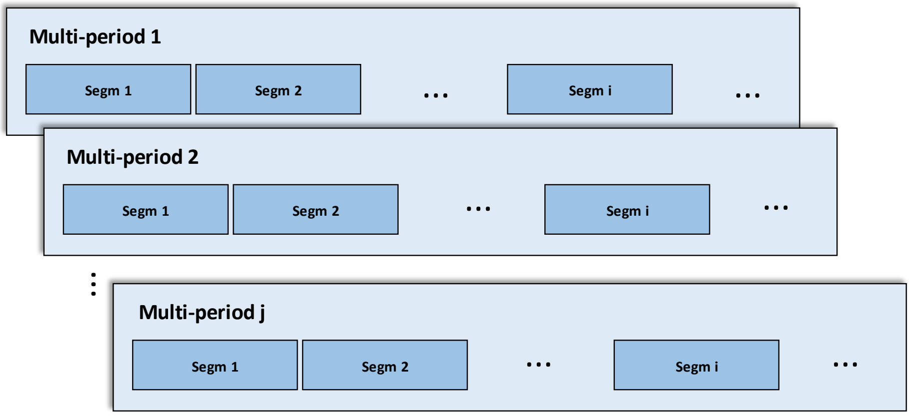


Figure 3.7.: **The multi-periods with segments.**

In the current work, it is assumed that a certain set of events have the possibility to appear over a certain time interval, and the term *weight* is adopted to convey this possibility. Suppose there are n segments in each basic period and the splitting is carried out in the same way, i.e., to slit all basic periods at same time points. The following form is used to denote event distribution patterns:

$$[I : S : W]$$

where I is the event set, S stands for the segment ID, and W is the weight of I w.r.t. S. Note that I, S, and W are multi-dimension. This provides the information of the event set I has weight W in segment S.

On the one hand, an event set may have variant *weights* over different segments; on the other hand, one segment may also hold several event sets. In a segment, the event set with *weight* over a predefined threshold is regarded as *evident*, implying that which event sets frequently occurs in which time intervals. As shown in Figure 3.6, if each of the basic periods of a multi-period P_{mul} are split into m segments in the same way, which means they have the same split time points, then say the multi-period also has m segments (see Figure 3.7). Thus, multi-periods P_{mul} can be described by m evident event sets. For multi-period P^j_{mul} with m segments, its profile has the following form,

$$P^j_{mul} := \bigcup_{i\in[1,m]} [I_i : S_i : W_i] \tag{3.4}$$

For a long term observation, many multi-periods make up the time-line. Thus the long term human behavior patterns can be profiled by using

$$Profi_{Beha} := \bigcup_{j\in(1,sum)} P^j_{mul} \tag{3.5}$$

where sum is the sum of multi-periods within the observation time-line.

3.4. Chapter Summary

Towards formally description of the in-home scenario, this chapter has proposed the abstractions of the in-home behavioral scenario. Firstly, the living environment was mapped onto the *@home graph*, a graph-based model. Considering the interaction between the human and the environments, the manifestation is abstracted as events, and considering the characteristics of simple sensors, two kinds of events have been categorized, i.e., *time-point-based events* and *time-interval-based events*. Finally, the event distribution over time intervals has been regarded as the indicators of the human behavior pattern where a hierarchical form of patterns abstraction is proposed.

Chapter 4

Human Behavior Profiling

As known, in the context of healthcare with pervasive computing, the resident's behavior information is hiding in the dataset of event. Considering healthcare application, the purpose of building a sensor-enhanced environment is to provide more valuable information. Based on the abstraction of in-home scenarios, which has been introduced in the previous chapter, this chapter will be devoted to the issue of information discovery from the dataset, which is measured through pervasive healthcare. Meanwhile, some content is introduced by connecting to the GAL-NATARS study, a study designed for promoting continuous observing.

At the beginning, section 4.1 firstly address the data format measured through some common simple and unobtrusive sensors, and further gives the methods to extract events from raw data. In section 4.2, the methods for extracting the information of lifestyle pattern from event dataset will be proposed. Several metrics are defined in section 4.3 so as to profile human behavior from some aspects.

4.1. Understanding the Data

To implement pervasive computing for practical application, unobtrusive sensors are relied upon to gather data. Here the unobtrusive sensor refers to such that has hardly influence on resident's normal life and the resident can freely do activities as usual. In other words, the sensor are acceptable and it brings no extra burden on normal life. Though the data gathered by these sensors has simple format, it still has its own characteristics. Considering the operation mechanisms of sensors, which have been introduced in section 2.4.4, towards better understanding the data, some preprocessing has to be carried out on the data before it can be used for extracting information. This section firstly puts forwards the data format recorded by these sensors, and then proposes the preprocessing method, referring to event extraction.

Some example data gathered in the GAL-NATARS study is used in this section for convenient explanation.

4.1.1. Data Format

For the purpose of detecting parameters in demand, multiple sensors are usually chosen to build a pervasive in-home healthcare system, thus generating a dataset with various data formats. Based on the signal reporting mechanism, the sensors used can be classified into two categories: one includes the sensors that are triggered by some specific situations, named as *passive sensor* in current work, such as contact sensor and motion sensor, and the other category of sensors work based on sampling the signal in fixed frequency, named as *active sensor*, such as the electronic current sensor and the accelerometer.

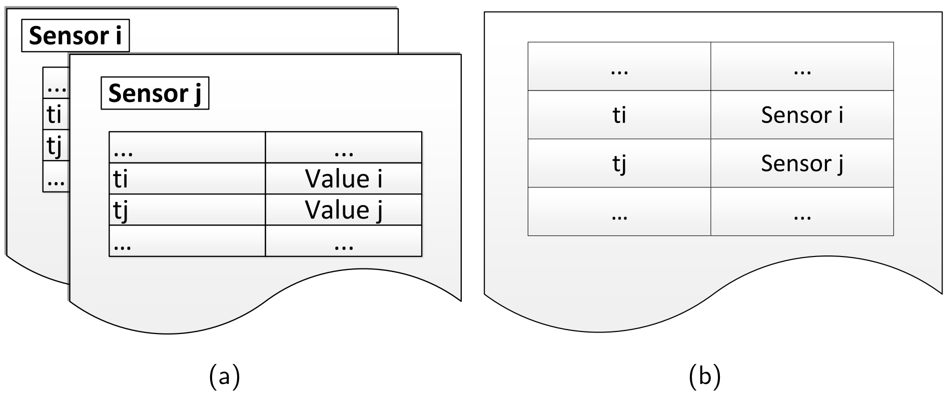


Figure 4.1.: Data formats of passive and active sensor.

For *passive sensors*, the data format simply consists of time stamp and sensor identifier, indicating when the sensor is triggered and where or which sensor is triggered respectively (see Figure 4.1 (a)). However, because of the variance of sensing mechanisms and different interactions between various objects and the resident, the distributions of records over time for different types of sensors represent distinct performances. To demonstrate the sensor record distribution, some example records from the real-life observation are presented in Figure 4.2. With respect to the contact sensor (reed switch sensor), each record of the sensor means one time contact of two objects. For the application in the GAL-NATARS study, the events of door opening and door closing trigger the sensor. Normally, the records for the manipulation on the door should occur in pair, referring to the door opening and closing, and the interval of these pairs of records should be relatively short (the intervals are no more than 50s in Figure 4.2). With respect to the vibration sensor, any action on the object detected that generates vibration with higher amplitude than the predefined threshold will generate one record. Because the sensor effects on the specific object, and in addition, as a result of the setting of sensor's idle time, the records distribution is in accord with the frequency of the actions on the object. Because the vibration sensor and the contact sensor are triggered only if the resident contacts with some objects, the records of vibration sensor are usually relatively sparse. Yet as to the PIR sensor (motion sensor), the records distribution is along with the distribution of the movements in sensing area. The density of records might be quite high if the inhabitant has high activity level, otherwise, it also might be quite low. Figure 4.2 shows that mostly the records of PIR sensor occur with high density, while there are still certain proportions with long intervals.

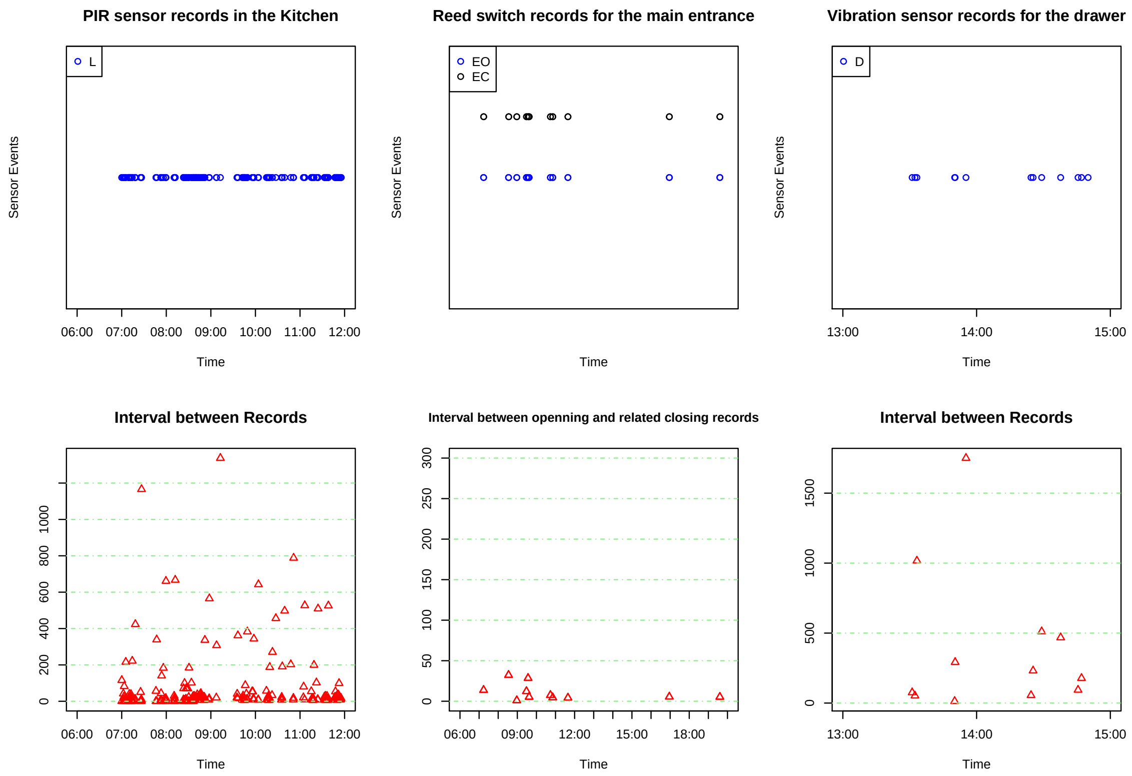


Figure 4.2.: The example records of three kinds of sensors. The first column demonstrates the records of PIR sensor in the living room and the interval between them; the second column demonstrates the records of reed switch mounted on the main entrance of the apartment; the third column demonstrates the records of the vibration sensor deployed at the cutlery drawer. L: Living room; EO: Entrance opening, EC: Entrance closing; D: cutlery Drawer.

For *active sensor*, since signal is sampled and reported in a fixed frequency, the data format consists of three components, i.e., sensor identifier, time stamp, and sensing value (see Figure 4.1 (b)). To avoid information missing, the sampling frequency

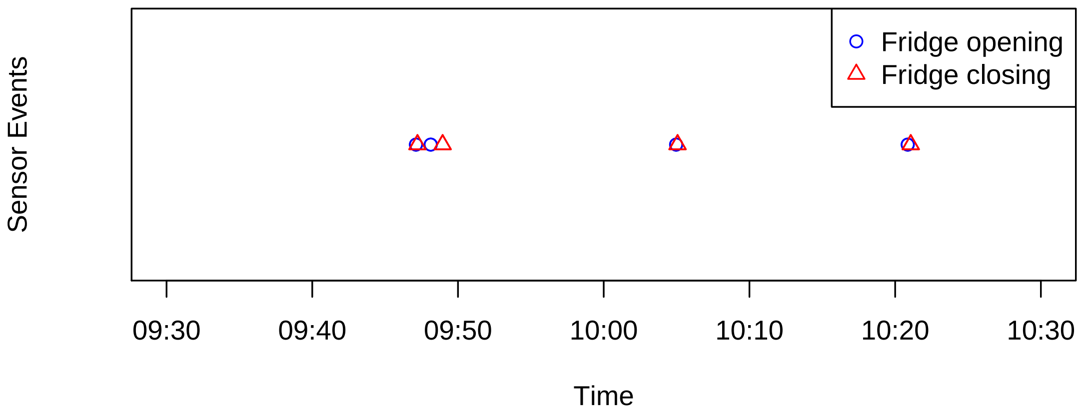


Figure 4.3.: Example of reed switch sensor records on the fridge.

is usually much higher than the frequency of occurring events, which means more data is collected with high density, resulting in much more information redundancy. As shown in Figure 4.4, the TV set is measured by the electric current sensor and its power value is recorded every second. From the plot it is evident that there are many sensor records between the time interval, from 09:30 to 10:30, but what is of interest is only the change points that can show when the state of this device change, i.e., from *off* to *on* or otherwise. As for Figure 4.4 this point is around 09:56. Therefore, regarding this type of sensors, first of all, the change points have to be detected to extract useful information, whereby the data density is dramatically reduced, and what's more, the data formats of both types of sensors can be unified after this pre-processing. In the current work, a sliding window is used to detect the event of state change for active sensor. As shown in Figure 4.5, the sliding window consists of two sub-windows with the same size. The stepwise width is set to be the same as the size of sub-window. Based on rated power of devices that are measured, the mean values of two sub-windows are used to identify state change. If a dramatic change between mean values of two sub-windows is detected, say there is state changes between two sub-windows. More precisely, if $mean_{sub1} < th_{active}$ and $mean_{sub2} > th_{active}$, then say the state changes from *off* to *on* at middle of sliding window. Otherwise, if $mean_{sub1} > th_{active}$ and $mean_{sub2} < th_{active}$, then say the state changes from *on* to *off*. In the case of plugging out the power cable of a device, the real time distance between endpoints of the sub-windows also has to be taken into account. Based on the unified data format, the extraction of events can be carried out in uniformed methods.

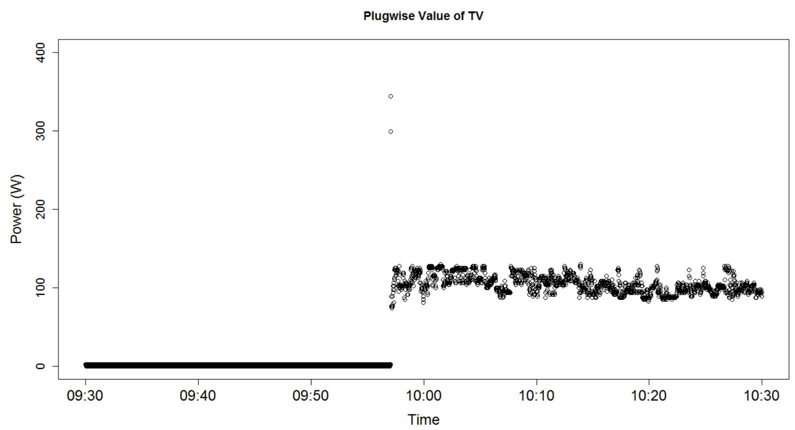


Figure 4.4.: **The power value of TV measured by current sensor.**

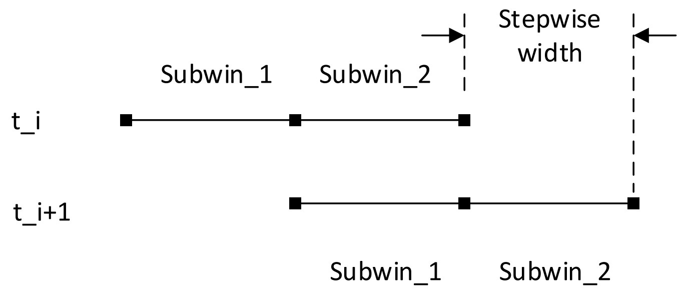


Figure 4.5.: **The sliding window used to detect state change event.**

4.1.2. Extraction of Events

As introduced in chapter 3, the *state* change of components in *@home graph* is abstracted as *events*. The first issue of information discovery is how to get *events* from raw datasets, i.e., *event extraction*. Event extraction refers to determining the state of a component in the *@home graph*, more accurately, determining when the state is *active*. The process of event extraction of one component is mainly on the basis of the sensor used to detect the component and the topology of the *@home graph*.

According to the abstractions proposed in chapter 3, the events of sensors are grouped into two categories, i.e., *time-point-based event* and *time-period-based event*. In line with the previous introduction, the following sections will also introduce the methods of event extraction in two categories.

4.1.2.1. Extraction Time-point-based Events

According to the definition, the *time-point-based event* $E[e_P : t]$, where $t :=< t.o >$, can be described as: the event e_P occurs at time $t.o$. For components detected in

@home graph, the sensor ID shows which component is detected, and time attribute, the occurring time t_o, has to be identified, so as to show when the component is *active*. For instance, for the vibration sensor used in the GAL-NATARS study, the operation on drawer should be dealt with as time-point-based event which satisfy the form $[e_P : t]$. As introduced above, the reed switch sensor is used to detect the manipulation with some often-used doors, such as the main entrance of the apartment and the door of the fridge. Intuitively, it is inferred that this kind of events should occur in pair, which can be confirmed through the data records (see Figure 4.3 as an example). Hence, with respect to this kind of records, either time-point-based or time-interval-based events can be extracted. If the door opening and closing are regarded in pair, then it can be described as: the duration of the event *Door d opening* starts at time $t.s$ and ends at $t.e$, whereas if they are considered separately, then there are two events occurring, i.e., *door d opening* at t_s and *door d closing* at $t.e$.

4.1.2.2. Extraction of Time-interval-based Events

According to the definition, the *time-interval-based event* $E[e_I : t]$, where $t :=< t.s, t.e >$, can be described as: the event e_I starts at time $t.s$ and ends at time t_e. Similar to *time-point-based events*, both operation mechanism of corresponding sensor and the topology of the *@home graph* have to be taken into account when extracting events. Moreover, since interval-based events may be either on one node or on a group of nodes, they have to be regarded separately in both situations.

According to the sensing mechanisms of the unobtrusive sensors, it is expected that the PIR sensor can be used to extract the time-interval-based event. For understanding the data, Figure 4.6 shows a section of PIR sensor records as an example. Recalling the introduction of mechanism of PIR sensor, a sleep time T_s is set to the PIR module and there is a minimal interval between two consecutive records. However, because the PIR sensor detects motions inside sensing area, the intervals are not always equal to the minimal sleep time T_s even when the resident stays in the sensing area during that interval. Due to sensor sensitivity and deployment, there are considerable intervals bigger than T_s even in the time interval with high record density in Figure 4.6, such as the events around 14:30. Therefore, if the interval-based event is wished to be extracted through the PIR sensor, the position of the resident has to be located by checking the states of other sensors.

In the case that a living environment is fully covered by the motion sensors, there is no chance for no sensor triggered as long as the resident has movements that are intense enough to trigger the sensor in the apartment. Note that if the resident transfers from one sensing area to another, there is a short interval between the two sensors related. In order to extract the interval-based events based on the records

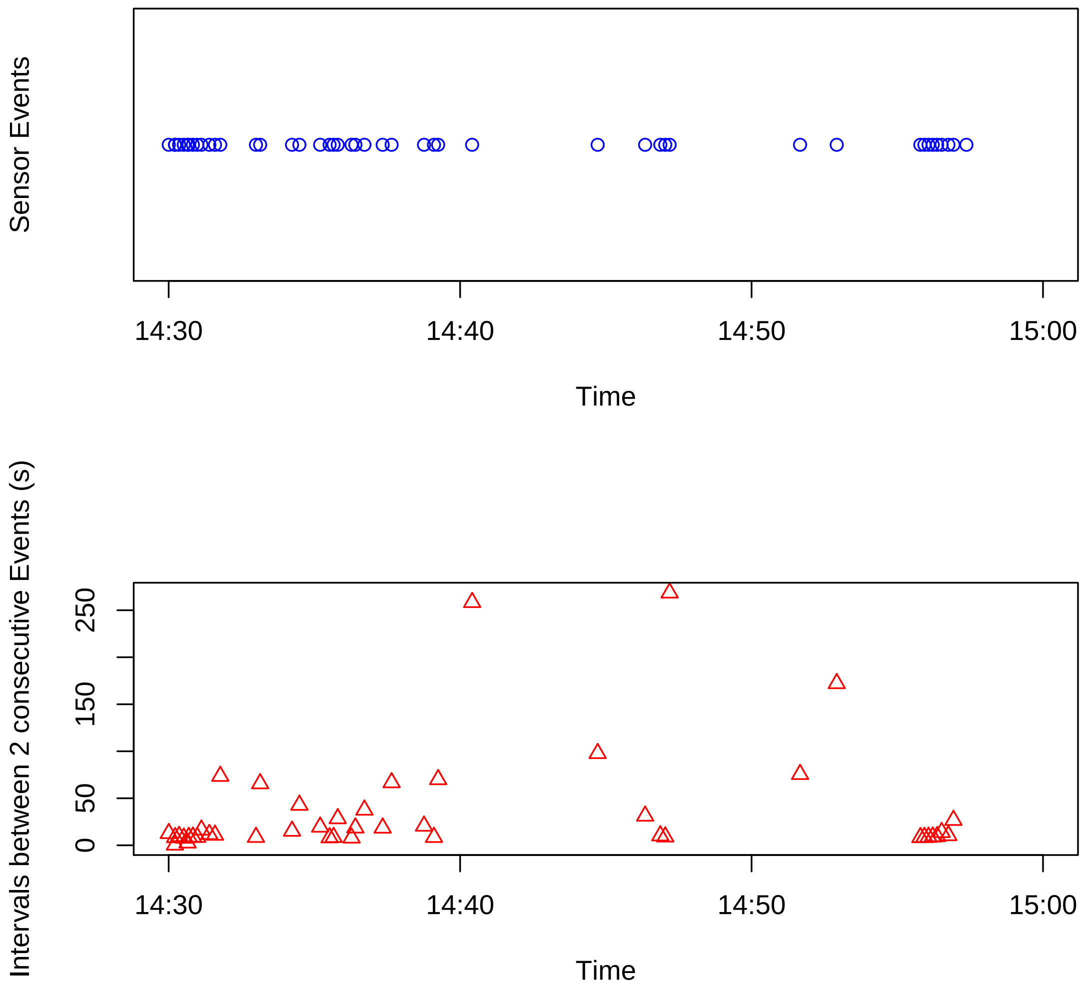


Figure 4.6.: The example PIR sensor events detected in the kitchen.

of the motion sensor, some criteria are made based on the abstractions of in-home behavior scenarios.

4.1.2.3. Extraction Criteria

Considering the data format and the topology of the *@home graph*, a number of extraction criteria are made.

Suppose only one resident lives in the living environment. The form $C_i \vDash S_j$ is used to denote that the component C_i is detected by sensor S_j. The term *active* and *inactive* is used to denote the state of component in the *@home graph*, and the theorem 1 to determine the state of a group. Some notations are used in the definitions of next theorems: C: a component of *@home graph*; G: a group of components. To determine the state of a group, Theorem 2 is described as:

Theorem 2 *if* $C_i \in G_k \wedge state(C_i) := active$*, then* $state(G_k) := active$*.*

To determine the starting and ending of *active* state of a group, Theorem 3 is described as:

Theorem 3 *if* $state(G_i^t) := active \ \wedge \ state(G_j^t) := inactive \ \wedge \ state(G_i^{t+1}) := inactive \ \wedge \ state(G_j^{t+1}) := active$*, then* $start(G_j) = t + 1, end(G_i) = t$*.*

The starting of a event is always follow the ending of another event, otherwise, the ending is before the next starting. Considering the sensor sensitivity, the starting is more sensitive than the ending, here the starting of next event is used to decide the ending of the previous one. If the current backbone node is inactive and no other backbone node is active (e.g., detecting for the other room occupying), then the resident is treated as still staying in the current sensing area (current room). Given the scenario that the resident, who is living alone, is sitting in the living room, and right now he or she is performing no obvious movement (see Figure 4.7). The PIR sensor cannot detect anything in this situation, but there is some records about the person since he or she have surely triggered the sensor when he or she enters the sensing area. That means since the person keeps stillness, there is no record made by the PIR sensor for the living room. On the other hand, the PIR sensor for the hallway, which is a neighbor backbone node of the current backbone node, also detects nothing. In this case, it is implied that the person is still staying in the living room, and the event continues until the PIR sensor for the hallway is triggered.

Theorem 4 $C_i \in G_k$*,* $state(G_k) := active$*,* $t \in [start(G_k), end(G_k)]$*,*

* *if* $\exists T_s, \forall t < T_s, state(C_i) := inactive$*, then* $start(C_i) = T_s$
* *if* $\exists T_e, \forall t > T_e, state(C_i) := inactive$*, then* $end(C_i) = T_e$

Apparently, the events of leaf nodes are always *during* the events of the group they belong to. For simplifying the extraction, as to the time-interval-based events on leaf node, the first record of corresponding sensor is regarded as the starting mark and the last record as the ending mark. The extraction criterion is formulated as theorem 4.

Figure 4.7.: An example apartment layout.

4.1.3. Detection of Multiple Persons

According to the assumption, the living environments in the current work are constrained to the ones with only one resident. The study inclusion criteria of the GALNATARS study also states that only the older persons living alone are recruited as the subjects. However, it is still possible that some visitors come to the apartment at times.

As known, the target of building a sensor-enhanced environment is to evaluate the resident's health statue and his or her performances. If there are more than one persons appearing in the living environment, the sensor records triggered by those persons will be mixed together, and it is difficult to separate the target person's data or to identify which resident's patterns are reflected. Therefore, with the assumption that the resident lives alone, *detection of multiple persons* will be carried out to show the visiting situations.

According to chapter 3, the living environment can be represented by using the *@home graph*, through which the topology is clearly shown. Intuitively, the most different performances between one resident and multiple residents is the time interval and the distribution on the graph of consecutive records. If there is only one resident, sensors will be triggered consecutively with reasonable time distances; otherwise, if there are more than one person and they appear in different positions, e.g., in two rooms, some sensor records may be triggered within very short time, and in other words, there are sensors triggered nearly simultaneously, which cannot happen in the situation of one resident.

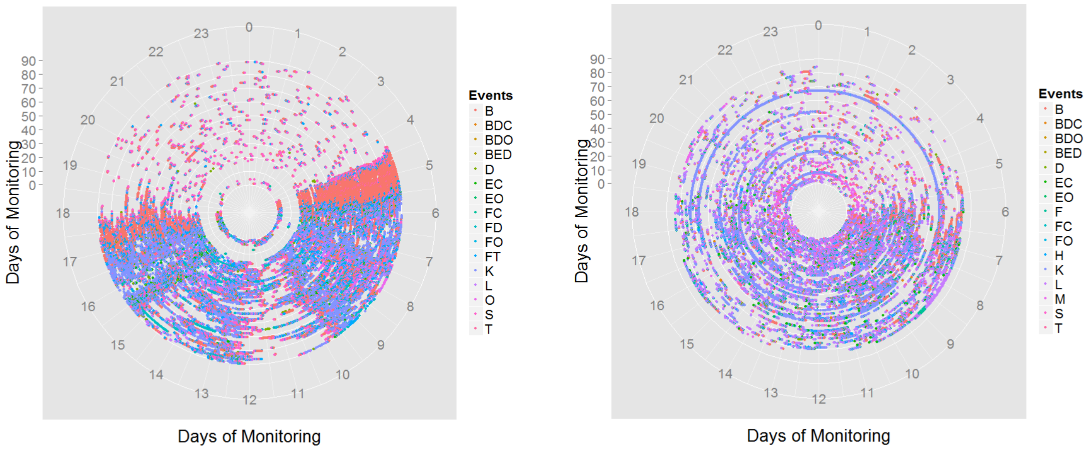
Towards detecting multiple persons, the concepts of *jump* and *fast transfer* are defined. Suppose there are sensor record s_i from *group m* and s_{i+1} from *group n*. If *group n* is not the neighbor group of *group m*, and the time gap between them is shorter than a threshold th_j, then say there is a *jump* between these two records; if *group n* is the *neighbor group* of *group n* and the time gap between them is shorter than th_f, then say there is a *fast transfer* between these two sensor records. A *jump* means there are at least two persons in the living environment. A *fast transfer* means there are probably more persons during the corresponding time interval, and this data of that segment is not reliable. Both *jump* and *transfer* imply the data around them are unreliable.

4.2. Temporal Pattern Discovery

During daily life people perform a range of activities that can be manifested by the events occurring on abstraction model, which is *@home graph*. For the purpose of comparing peoples' activity distribution over time line, two datasets that are measured from two subjects' living environments are presented by means of spiral plot (see Figure 4.8 [103, 104]). In spiral plot, the radius direction means the days of monitoring, from the first day in the inner circle to the last day in the outer circle, and each day's records are plotted in one circle. The circle direction (clockwise) indicates the clock per day. The colored points ordered in spiral denote the sensor records.

From these two plots it is evident that these two residents have exhibited quite different lifestyles in terms of sensor records distribution. Obviously, the right side one seems much more restless during night than the left one, and the left one has significant boarder between nighttime and daytime, and some events occur frequently around 5:00 and 18:00. There is much more difference between daytime and nighttime in the left one, while the situation is not like this as to the right one. Therefore, it is implied that:

Different individuals can exhibit different impressions through data distribution.



(a) (b)

Figure 4.8.: Spiral plots of two subjects' data records.

From the previous section it is already known that events can be extracted from raw datasets. Based on the extraction of events, Figure 4.9 presents the events of a number of consecutive days, where the x-axis is the clock of one monitoring day and the events are denoted by points (for time-point-based events) and segments (for time-interval-based events). From Figure 4.9 shows that during nighttime (from 00:00 to 06:00 and from 22:00 to 00:00) there are some events emerging, but fewer than daytime. In addition, it appears that at night the resident spent most of his or her time in the living room rather than the bedroom as normal people (see the red segments). In general, even though variance exists, the resident's habits are reflected in the form of the event distribution. Therefore, it is implied that:

Across different time intervals the same person exhibits different event distribution.

To sum up, on the one hand, due to the variance in cultures, health status, characters, and habits, different individuals can exhibit various impressions of living patterns (or lifestyles); on the other hand, people usually experience periodic lifestyle. All specific patterns are connecting to the time factor, i.e., patterns are represented by time distribution. Consequently, the discovery of temporal patterns would be helpful to interpret event datasets, so as to extract useful information. Combining with the abstraction presented in chapter 3, this section proposes the methods of temporal pattern discovery. In particular, this section will answer the question: *How to depict the resident's lifestyle?*, which can also be stated as:

Which events usually arise in a certain time interval?

or

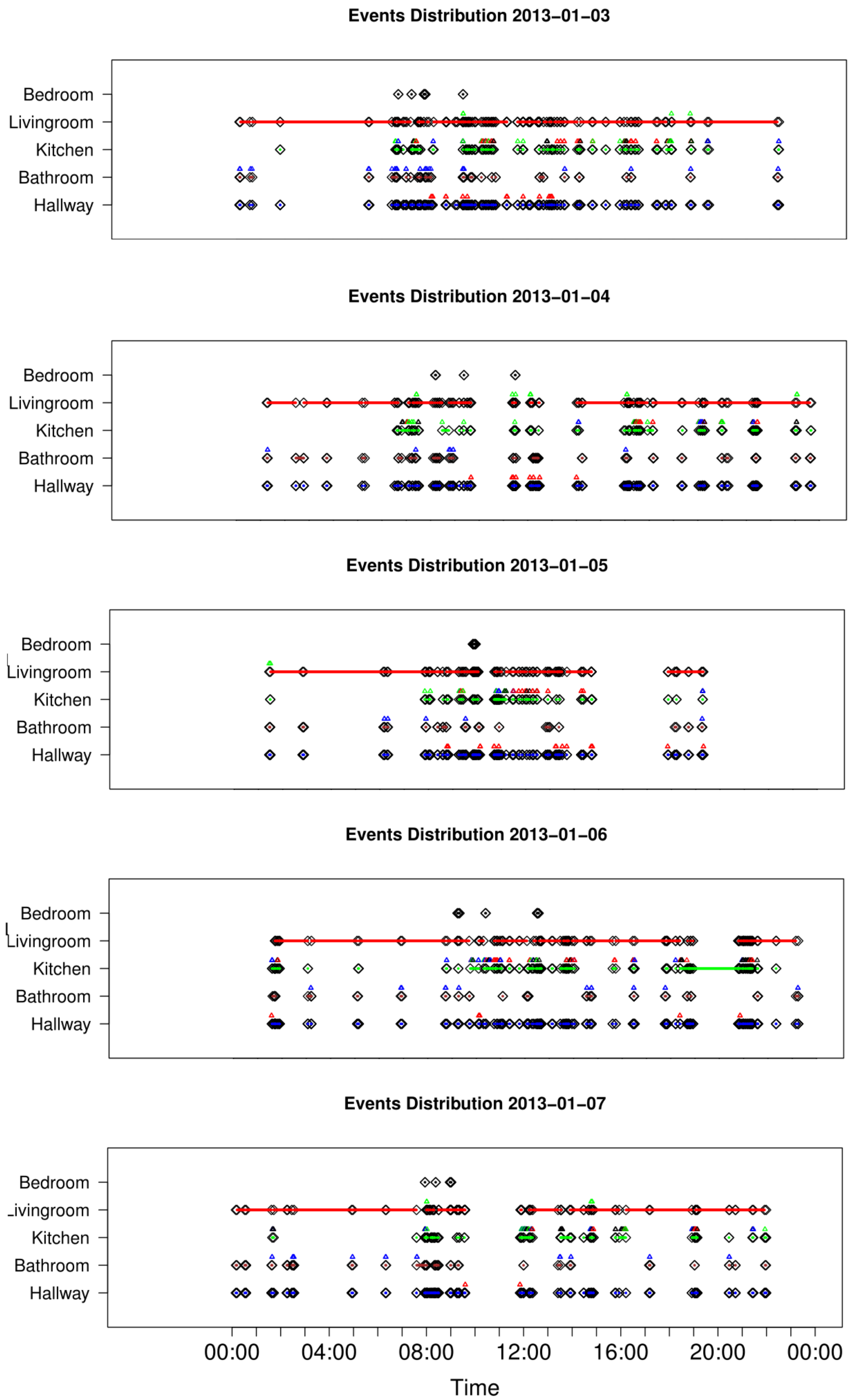


Figure 4.9.: A number of consecutive days events of one resident.

The resident usually does ... (what) ... (where) ... (when).

4.2.1. Data Reorganization

The extraction of events converts the dataset in the form of record sequence into the form of event sequence. Based on the assumption that the regularity of lifestyle is an important indicator of health status. This section will introduce the method to reorganize the dataset.

4.2.1.1. Time-line Splitting

People's daily behavior distributes over time-line, and normally some specific activities are performed within some particular time intervals. For instance, a resident usually cooks in the morning, midday and evening, and there is not likely to be any cooking activities at late night; having a walk out of the apartment may happen in any time of the daytime, except for nighttime. Therefore, in this case the time-line of one day (24-h) can be split into several segments taking the normal people's habits into account. Similarly, for a person who is not retired, during weekdays there tend to be fewer activities in the daytime at home, while more events might be detected during weekend. For a person with active lifestyle he or she might do some daily sports in similar time, while there might be much more time spent in watching TV or surfing Internet for a person with inactive lifestyle. Consequently, people maintain some regular habits, and meanwhile various lifestyles are represented with respect to different individuals.

Another issue having to be considered is fluctuation in human habits. Despite that people periodically perform some activities, the daily living can be influenced by some uncertain causes, thus resulting in non-strict lifestyle. In this context, the *drift* of events occurring time is possible. Consider the following case. A resident usually gets up at 7:00, but he or she may get up some time later due to the sleepy condition or even the bad weather, otherwise he or she may get up some time earlier. So in order to analyze the *drift* of human lifestyle, a hierarchical segmenting is applied to each basic period P_b. Suppose the basic period is set to be 24-hour. As shown in Figure 4.10, based on normal people's habits, the time-line of 24-hour is split into a number of segments, and further some consecutive segments can also construct long-segment. In Figure 4.10, the red bars denote the long segments which are named as “late night”, “morning”, “noon”, “afternoon”, and “evening and night”. Each long-segment contains two or more segments, which are denoted by black bars in the bottom.

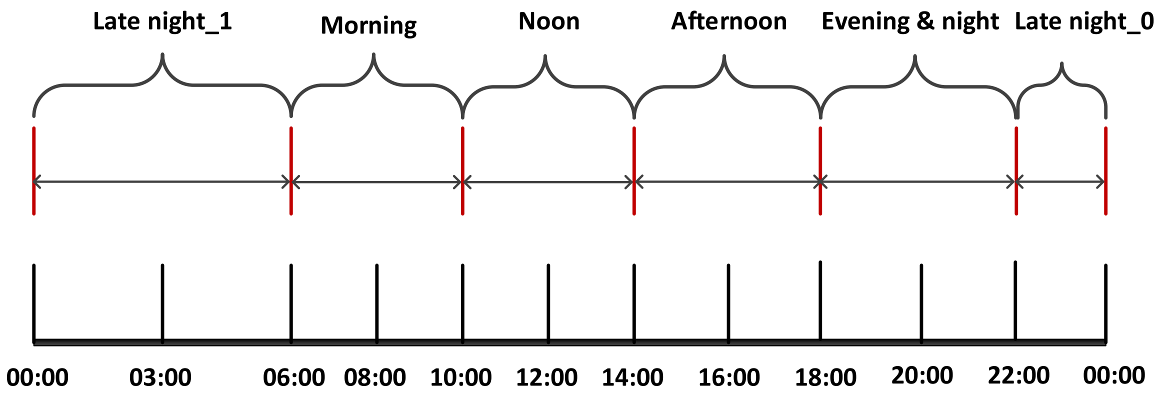


Figure 4.10.: **The segmenting of one day (24-h) time-line.**

4.2.1.2. Event Sequence Clustering

The events on backbone nodes and leaf nodes constitute a sequence that is organized in time order. One key issue having to be addressed is how to reorganize the sequence into more structured style.

People's activities are usually performed in groups. In other words, the activities occur in a kind of rhythm. A certain set of activities usually occur following others. For instance, if the resident enters the living room at night, he or she tends to turn on the lamp and turn on the television, while if he or she is in the kitchen, the oven is likely to be switched on and the door of the fridge may also be opened and closed. If he or she gets up at night, the bathroom and toilet tend to be used. As soon as a set of activities occur, the time distances between them are short. People usually carry out a series of activities one after another, and then stay still for a while. As manifestation of activities, the distribution of events is also rhythmic.

To highlight the rhythm reflected by human behavior, the sequence of events is clustered by splitting it into episodes. In this process, the splitting constraint is *time gap*, which is defined as the time distance between two consecutive events. More formally, the splitting criterion is stated as theorem 5.

Theorem 5 *Splitting of the sequence of events: Suppose the time gap between two events is* $gap(e_i, e_j)$ *, if* $gap(e_i, e_j) > th$*, where th is the threshold that a value can be set as to application, then this two events should be split into two episodes.*

As known, the time attribute of time-interval-based events includes the *starting* and the *ending*. The span of a time-interval-based event can be various. For those long activities, which are usually extracted as time-interval-based events, they are actually resident's state, e.g., watching TV, reading newspaper, and sleeping. When

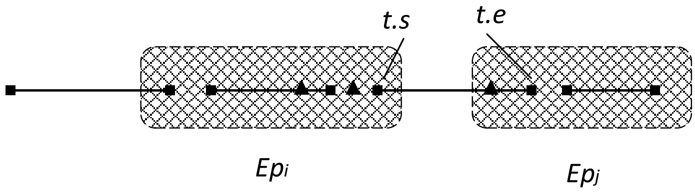


Figure 4.11.: Dealing with time-interval-based events.

the event is mentioned, what is of interested is the change of event. Therefore, time-interval-based event is regarded as two events in the process of splitting the sequence of events, i.e., *event starting* and *event ending*. As shown in Figure 4.11, the starting and ending of time-interval-based event e are included in episode Ep_i and Ep_j respectively.

Generally, the order of doing some activities is an important indicator of judging whether a certain task is successfully accomplished. In the context of mental impairment, order is usually used to assess the patient's rehabilitation. However, in the current work, the time order is not taken into account due to the following reasons. First, if the event order that is able to reflect resident's mental status is of interest, the density of deployed sensor should be high enough to detect the events needed. This issue is beyond the dataset. Second, according to the process of splitting sequence to episodes, the time constraint is able to represent that the events included in one episode have higher correlation, and the variance in their time order cannot tell difference in health status.

Now that the time order is not taken into account for those events in the same episode, it seems that the events in the same episode are loaded into a *bag*, and they are regarded as they occur in the same time interval and their order is omitted. Consequently, the sequence of one period's events is converted into a number of *bags* of events. Even though there is no time order within a bag, there is still a time span for a bag of events. Considering the time-line splitting, with regard to the bags whose time span covers two time segments, which segment the bag should belong to has to be decided. In the current work the majority priority is adopted. Suppose there are two consecutive time segments s_i and s_{i+1} and an bags of events with time span from $b.s$ to $b.e$, if more than $(b.e - b.s)/2$ falls in s_i, then this bag belongs to s_i, otherwise, it belongs to s_{i+1}.

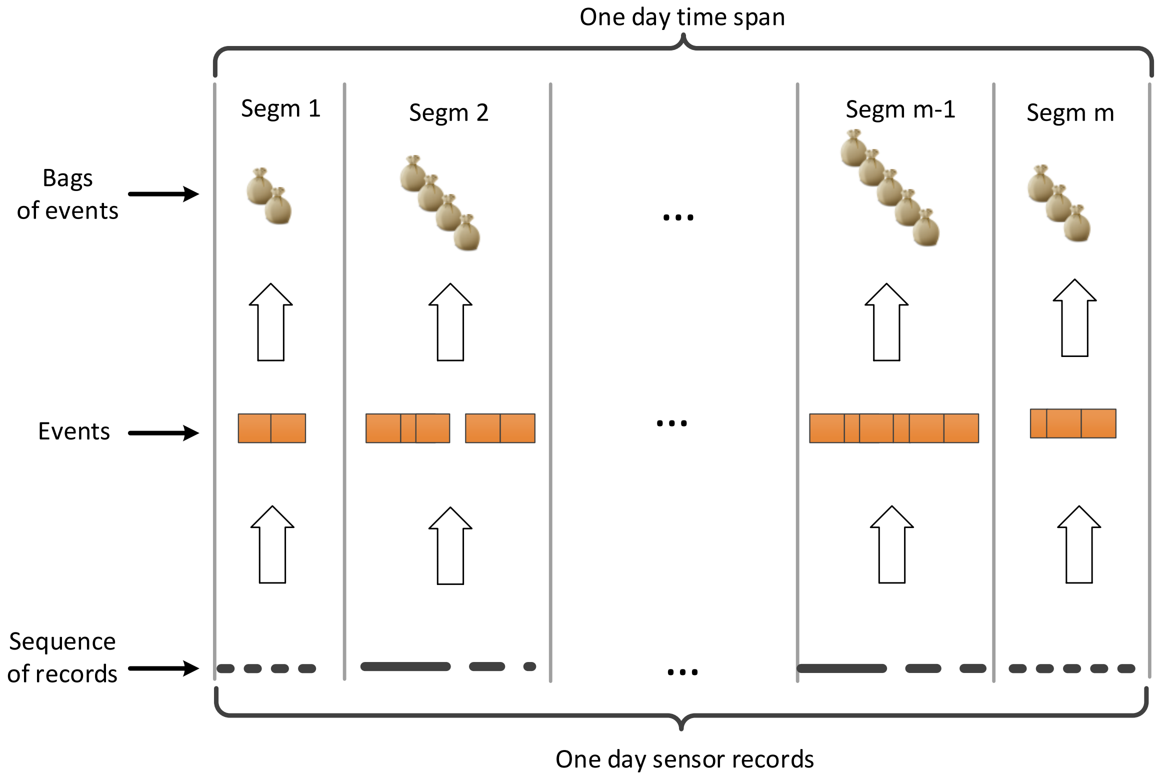


Figure 4.12.: The procedure of data preprocessing.

To sum up, the procedure of data reorganization is firstly clustering event sequence into episodes on the basis of event extraction, and then loading those episodes of events into bags, and meanwhile the bags are restricted in the segments of time-line. After this processing of raw data, the whole dataset will be represented in more structured form The procedure is illustrated in Figure 4.12.

4.2.2. Discovery of Frequent Event Sets

As what is previously explained, a feature of living environment is that events occur in groups, which means some events tend to appear as a cluster because of the rhythmic activities performance. Within a time interval, the most frequent event set implies that the resident devotes to corresponding activities in that time interval. It is reasonable to leverage these event sets to indicate human lifestyle. In order to fill out the statement raised at the beginning of this section, a systematic methods will be proposed to discover the temporal pattern based on association rules discovery.

As introduced in chapter 2, a method of implementing association rules discovery is *apriori algorithm*. As shown in Figure 4.13, the input of apriori algorithm are a dataset in the form of *transactions*, and some arguments such as minimum support and minimum confidence. For specific implementations of the algorithm, there are also other arguments, such as the minimum length and the maximum length.

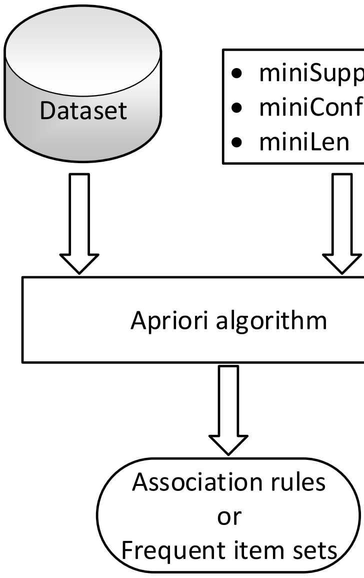


Figure 4.13.: The usage of apriori algorithm.

After event clustering, the original dataset has been converted into bags of events which is similar to the dataset of *transactions*, the standard data form of apriori algorithm. However, to highlight the frequent event set within a certain time interval, the input data have to be reconstructed. Normally there are much more events during daytime, while much less at night. If the whole day's data is used as the input, the events happening at night cannot be shown due to their low density comparing to the daytime. Similarly, some typical events hidden in special time intervals are also covered by the others and cannot be discovered. Moreover, even the events with higher probability w.r.t. a certain day are known, but more detailed time information is still unknown, such as during which time interval they occur. So the statement still cannot be filled out. Therefore, with regard to the current application context, an input construction mechanism is firstly proposed in this section; Afterwards, the method of selecting results is also formulated.

As shown in Figure 4.14, suppose there are n consecutive periods that are selected by a sliding window in one time's analysis. For clear illustration, these n periods of data are placed in parallel. Then they can be regarded as a multi-period consisting with n basic periods. Because each basic period is split into m segments, the multi-period can also be regarded as having m segments. According to section 3.3.1, if there is an event set, E_i , and it occurs in s^{th} segments of basic periods P_i and P_{i+1}, then say E_i is *periodic* as to these two periods. Because of the fluctuation existing in the regularity of human lifestyle, even if an event set is periodic, it can still appear at any time point within a segment of a basic period. For *segment 1* shown in Figure 4.14, an event set (the black dot) occurs in the middle of *segment 1* in P_1, while close to the end of *segment 1* in P_2, and close to the start in P_n. To address the fluctuation, the segment is further split into s subsegments. Note that s may vary corresponding to different segments, and there are usually more subsegments in segments with high event density than in those with low density. In a multi-period, the segments in the same position of each basic period make up a column, which is named as *segment column*. Obviously, there are a total of m *segment columns* in one multi-period. Since each segment is additionally split into s subsegments, a *segment column* is actually a matrix with size $s \times n$.

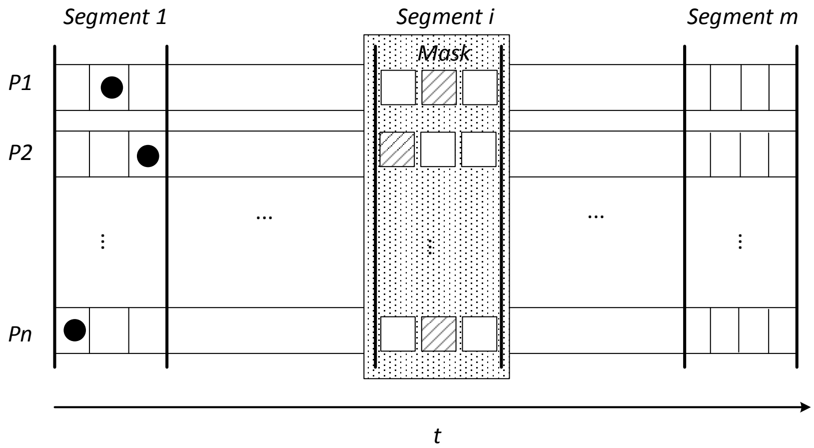


Figure 4.14.: A multi-period showing the construction of input data.

In the next step a *mask* in the form of matrix is used to construct input data by selecting the bags of events. If frequent patterns from the time intervals with the same position across n basic periods are wished to be discovered, the corresponding *segment column* is covered by a *mask* with the size $s \times n$, which is the same size as the *segment column*. As shown in Figure 4.14, *segment column i* is covered by the *mask*. Each cell denotes a subsegment, and a *cell ID* is assigned to each cell. The basic thought of constructing input data is selecting certain bags of events in each basic period, and combining all the selected results from n basic periods as the input data. The selecting process is based on that the time attribute limits in one subsegment of *segment i*. In order to traverse all possible combinations in *segment column i*, a number of iterations are carried out. In each iteration, only one cell is marked as selected in each row of the *mask*. The bags of events falling in marked subsegments are selected as the input set of the current iteration. In each iteration, a event set with the biggest confidence value is shown up. The biggest one of all iterations is considered as the *evident* event set. Meanwhile the selected *path* of the *evident* event set that shows the combination in the form of a sequence of *cell ID* is recorded as well.

As known, the time-line of the basic period is split into a number of segments. To avoid missing information, the time intervals around splitting time points should also be considered. Not only are the segments covered by the mask and the constructed input data for the algorithm, but also the time intervals around splitting points are applied by the mask. Say there is a splitting point, t_i. A time interval $[t_i - t_0, t_i + t_0]$ is regarded as the segment column and a mask is applied on it to construct input dataset.

By using a matrix mask, it is possible to find more precise distribution of a certain event set over a analyzed segment. More detailed lifestyle can be obtained. As discussed in section 4.2.1.1, since human cannot maintain a habit in very strict time constrain, it is possible to show the *drift*, more or less. With a matrix mask on a segment of the time-line, the drift of a habit in that segment can be revealed, as well as the evolution of the resident's lifestyle. Additionally, this method is able to contribute to the measurement of lifestyle regularity, which will be introduced in next section. Within the analyzed multi-period, if a frequent event set appears in one segment, and less variance of the cell ID in the mask, then it means more regular lifestyle, otherwise less regular lifestyle. Since the size of *mask* is $s \times n$, to count all possible combinations of subsegments, the number of iterations is s^n.

After applying the algorithm to the reconstructed datasets, the event sets with the support bigger than the minimum support sup_{min} are discovered. To extract the most useful information from the frequent events, a number of sets of interesting events are defined, referring to *interesting event sets* (IESs). Besides, a threshold th is set to select the discovered frequent event sets. If the support of a IES_i is bigger than th, say IES_i is *evident*. A subset of IESs appear in a time interval, either the segment or the interval around splitting time point. Thus a time interval can be characterized by its evident IESs. For multi-period P^j_{mul}, suppose there are m segments, and IES^i_{evi} is used to denote the evident IES in segment i, there is

$$P^j_{mul} := \bigcup_{i \in [1,m]} IES^i_{evi} \tag{4.1}$$

which represents the multi-period by the evident IESs. More formally, the whole process is described in pseudocode (see Algorithm 2).

Algorithm 2: **Discovery of evident IESs**

Input: *dataset*, *minSup*, *minConf*, *minSup*, *minLen*, *maxLen*, *IESs*, *th*
Output: IES_{evi} with *support* and *path*
for all segment columns **do**
  for all *path* **do**
    $input.apriori \leftarrow events\, on\, path_i$
    $item.freq \leftarrow apriori(input.apriori\, minSup\, miniLen\, maxLen\, IESs)$
    for all IESs **do**
      $c1 \leftarrow max.support(item.freq == IES_j)$
      $candidate_i \leftarrow \bigcup_i c1$
    end for
    $candidate_i$
  end for
  for all IESs **do**
    $c2 \leftarrow max.support(candidate_i == IES_k)$
    $IES_{evi} \leftarrow \bigcup_k c2$
  end for
end for

4.3. Metrics of Behavioral Profiling

According to the medical background introduced in section 2.2, the performances of some specific activities are able to indicate the human health status. Based on the methods proposed in the previous sections, a number of metrics are defined so as to profile human behaviors.

4.3.1. Measurement of Time Consumption

According to section 3.2, the time attribute of the time-interval-based event indicates the starting time and the ending time of a event. According to the definition of the time-interval-based event and the format of sensor data, entering and exiting a room can be extracted through the motion sensor data. The occupation of a room is regarded as time-interval-based event, and its time attribute is described as $< r.s, r.e >$, where $r.s$ and $r.e$ denote entering and exiting room r, respectively.

Suppose n time-interval-based events describe n times usage of room r of one day, i.e., the sequence $S_r = < r : t_i > | 1 \leqslant i \leqslant n$ denotes the usage of r. Therefore, the daily total time spent in room r, T_r, can be calculated through

$$t_r = \sum_{i=1}^{n} (r.e_i - r.s_i) \tag{4.2}$$

As to the frequency of a specific room or object usage, f_r, it can be easily calculated by counting the corresponding events occurring within that period.

4.3.2. Measurement of Activity Level

In the context of sensor-enhanced environment, the activity level can be obtained through a number of ways, such as counting the total number the triggering of sensors, counting the triggering of a specific sensor, and calculating measured value of a certain sensor. The living environment has been mapped onto the *@home graph* in chapter 3, and various topologies of the *@home graph*s are usually generated w.r.t. different living environments. Due to the difference in the position in the graph and the denoted object, some nodes are seldom active, while the others have more active states. In the current work, the backbone node at the *central* position of the *@home graph* is chosen to calculate the resident's activity level. Once there is transferring, the central node is more likely to be triggered. The more times triggering means the resident transfers more frequently, which also implies the resident has a higher activity level. As events have already been extracted from raw data, the number of events that passing thought the central node is calculated as the metric to represent activity level. One time's passing through the zone is defined as *entering* the zone and *exiting* it. Since the occupation of a zone is usually observed through a motion sensor (PIR) and the time-interval-based events are extracted, the *entering* event and the *exiting* event always occur in pair. The number of passing through the zone can be simply obtained by calculating the number of *entering* events or *exiting* events.

4.3.3. Measurement of Walking Speed

Measuring the walking speed is also an approach to assess the patients' mobility during the rehabilitation of the patients with mobility limitation. Some assessments measure how much time it takes to accomplish a certain task, such as Timed-Up and Go (TUG) and Short Physical Performance Battery (SPPB).

For the resident living alone, the sensor records form a sequence that represents the moving trajectory. According to the extraction of events, when the resident moves across some zones in the apartment, for example, the occupation of zones is a kind of time-interval-based events. The movements can be represented by a sequence of time-interval-based events. Consider the following case (see Figure 4.15). When a resident transfers from zone $Z1$ to $Z3$ through $Z2$, the corresponding sensor event sequence is { $Z1E$, $Z2S$, $Z2E$, $Z3S$ }, where the letter 'S' and 'E' denote *starting* and *ending* of a zone. For example, $Z1E$ means ending the zone $Z1$.

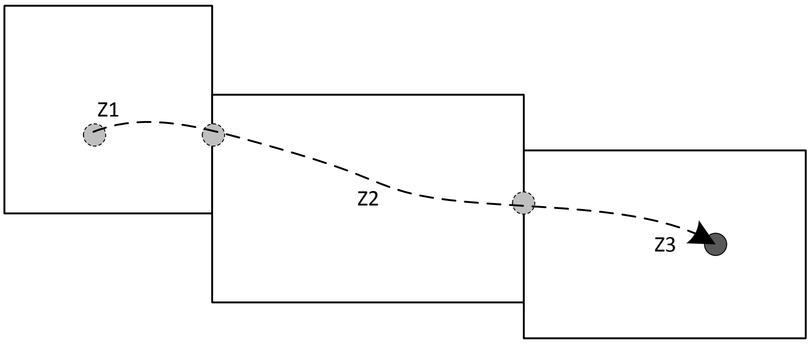


Figure 4.15.: Transferring of resident from $Z1$ to $Z3$ through $Z2$.

Obviously, zones are static. Theoretically, if one or more specific zones are chosen as the reference, then the resident's walking speed can be assessed. Since the *@home graph* ignores the distance and size of zones, the distance between two connectors is unknown. The change of time needed to accomplish the sequence is used to reveal the resident's mobility. For this case, if choose $Z2$ is chosen as the reference, the difference of time stamps of $Z2E$ and $Z2S$ can be used to indicate the resident's performance of passing through $Z2$. In other cases, the sequence can be longer if farther distance wished to be used as to assess the walking speed.

4.3.4. Measurement of Lifestyle Regularity

Across a number of multi-periods, the change of IES_{evi}, a subset of IES over a segment, is able to manifest the resident's ability to maintain a lifestyle or the change of behavior during corresponding time interval. Consider the following case. Suppose the multi-period is *week* and basic period is *day*. In week W_i, the evident event set is $\{Toilet, Bath, Kitchen\}$ towards segment seg_j, which stands for the time interval $[06:00, 08:00]$, while in week W_{i+1} it turns out to be $\{Bathroom, Kitchen, Livingroom\}$. In this case, it implies that the resident's morning habit has changed to some extent from W_i to W_{i+1}. The resident uses less toilet in week W_{i+1}, and the reason might be he or she uses it earlier or later, and might also due to some problems of the digestive system.

Before putting forward the methods to measure lifestyle regularity, the problem is formally stated as following. For the dataset $\mathcal{D}$, suppose the IESs is set $\mathcal{I} = \{i_1, i_2, ..., i_l\}$, where l is the length of $\mathcal{I}$, i.e., the sum of IESs. The amount of multi-periods is *sum*, and each multi-period, P^j_{mul}, has m segments with same splitting mode. When the human behavior is analyzed, some segments are usually combined into *intended time interval* (iti) and the indexes of segments are not necessary to be continuous. A set agr_k is used to denote intended time interval, iti_k, which aggregates a number of segments. As introduced in chapter 3, a multi-period is expressed in the form of $P_{mul} := \bigcup_{i\in[1,m]}[I_i : S_i : W_i]$. Here vector v_i denotes the IES_{evi} in seg_i, i.e., *distribution*. The values of elements in v_i are determined according to the definition of evident IESs, which is formulated as equation 4.3.

$$v_{(i,j)} = \begin{cases} 1 & if\ sup(i_j)\ \geq\ th \\ 0 & else \end{cases} \tag{4.3}$$

where $j \in [1, l]$. For different segments, v_i may exhibit variance appearance. IES^i_{evi} is evident in seg_i, while IES^j_{evi} is probably evident in seg_j, and IES^i_{evi} is not necessary to be equal to IES^j_{evi}. Basically speaking, the multi-period is characterized through the evident IESs over its segments, where the appearance is seen as attributes of multi-period. The lifestyle regularity is expected to be represented through the attributes of multi-periods.

To measure resident's lifestyle regularity, two factors are under consideration, i.e., *IESs density* and *change*. *IESs density* is defined as the ratio between the number of evident IESs in the given time interval and the number of maximum possible evident IESs in the given time interval. More formally, for iti_k, its IESs density is expressed as the formula 4.4.

$$D^{(k)}_{IES} = \frac{\sum_{i\in agr_k}\sum_{j\in[1,l]} v_{i,j}}{length(iti_k)\cdot l} \tag{4.4}$$

The density of IESs represents to what extent the resident has periodically performed a certain activity. The higher IESs density of an iti, the more periodic activities the resident shows up.

Change refers to how the behavior patterns vary between two consecutive multi-periods. To quantify the change, two factors are considered, i.e., the *IES density* and the *distance* between two distributions. Suppose there are two consecutive multi-periods, P^t_{mul} and P^{t+1}_{mul}. The distance between two intended time intervals iti^t_k in P^t_{mul} and iti^{t+1}_k in P^{t+1}_{mul} is $dis(v^{(k)}_t, v^k_{t+1})$. In the current work, the distance is calculated based on hamming distance, and it can be derived as following.

$$hamming.dis(v^{(k)}_t, v^{(k)}_{t+1}) = \sum_{i=1}^{l} cc_i \tag{4.5}$$

and according to the definition of hamming distance

$$cc_i = \begin{cases} 0 & if\ v_t^{(k,i)} = v_{t+1}^{(k,i)} \\ 1 & if\ v_t^{(k,i)} \neq v_{t+1}^{(k,i)} \end{cases} \tag{4.6}$$

Because the maximum hamming distance is $length(iti_k) \cdot l$ w.r.t. iti_k, the normalized distance is

$$dis(v_t^{(k)}, v_{t+1}^{(k)}) = \frac{hamming.dis(v_t^{(k)}, v_{t+1}^{(k)})}{length(iti_k) \cdot l} \tag{4.7}$$

The change between $v_t^{(k)}$ and $v_{t+1}^{(k)}$ is defined as

$$ch(v_t^{(k)}, v_{t+1}^{(k)}) = \frac{dis(v_t^{(k)}, v_{t+1}^{(k)})}{mean(D_{IES\ t}^{(k)}, D_{IES\ t+1}^{(k)})} \tag{4.8}$$

which means the change is the ratio between normalized hamming distance and the mean value of IESs density of two *iti*s in the given time interval. Note that when a resident's behaviors are analyzed, it is possible that not all time intervals are taken into account. For P_{mul}^k, the change w.r.t. all *iti*s can be combined together.

$$change = \frac{\sum_{i \in itis} ch_i}{num_{iti}} \tag{4.9}$$

where num_{iti} is the amount of *iti*s. Based on the definition of *change*, the variance of the residents' behavior is able to be quantified, thus manifesting residents' ability to maintain a regular lifestyle.

4.4. Chapter Summary

Based on the previous chapters, this chapter systematically proposes the methods of extracting information from datasets that are collected through sensor-enhanced living environments. Firstly, the data format was explored in section 4.1. Based on the sensor operation mechanism and the abstraction of living environments, two types of events, time-point-based event and time-interval-based event, were extracted. To fill out the statement of the resident's behavior status, section 4.2 has proposed the methods to discover behavioral temporal patterns, where the association rules discovery is employed. Through event extraction and temporal pattern discovery, some information connecting to behavior can be obtained. A number of metrics were defined in section 4.3, which are able to measure human behaviors.

Chapter 5

Results

In order to evaluate the effect, the theories and methods proposed in chapter 3 and chapter 4 are applied to the datasets obtained in the GAL-NATARS study, which has been introduced in section 2.4. Section 5.1 will firstly describe the settings of the experiment. The results will be presented in two parts. Firstly, based on event extraction, section 5.2 presents some basic behavior information, including time spent in specific rooms, frequency of a specific zone's usage, and activity level. Secondly, section 5.3 presents the results of subjects' behavior patterns by means of evident IES (Interesting Event Sets) distribution and lifestyle regularity. To evaluate the usability of the information discovered by those methods, section 5.4 analyzes the correlation between the results obtained through current work and the results gathered through assessments.

5.1. Settings of the Experiment

In the current work, *four* datasets are analyzed, which are gathered from four subjects who suffered from hip fractures and are currently living alone, having been discharged from hospital, as introduced in section 2.4. The dataset includes sensor data, which record sensor events occurring in subject's living environments, and results of medical assessment tests, which are obtained periodically by the nurses visiting over the duration of the monitoring term. In the following content, the dataset w.r.t. subject i is named in the form of *subject-i*. The map of subjects' ID from the GAL-NATARS study to the current work is listed in table 5.1. In the current work, all methods are implemented via the open source programming language, R, using version 3.0.0 [104].

Table 5.1.: Map of subject ID from the GAL-NATARS to the current work.

ID in the current work	ID in the GAL-NATARS study
Subject-1	NATARS20
Subject-2	NATARS21
Subject-3	NATARS30
Subject-4	NATARS31

Table 5.2.: The basic period time-line splitting

Time intervals	names
00:00 – 06:00	Late night 1
06:00 – 10:00	Morning
10:00 – 14:00	Noon
14:00 – 18:00	Afternoon
18:00 – 22:00	Evening & night
22:00 – 00:00	Late night 0

5.1.1. Settings of Event Extraction

According to sensor operation mechanisms and deployments, in the current work the data measured through PIR motion sensors is used to extract time-interval-based events; the data measured through the others, such as reed switch sensor, vibration sensor, and electric current sensor, is used to extract time-point-based events. For the state change detection of the electric current sensor, the threshold mean value is set to be 2/3 rated power value of the measured devices and the sliding window width is set to be 16, consisting of two sub-windows with size 8.

Based on apartment layouts, some relatively big areas are regarded as zones and they are mapped onto backbone nodes in the *@home graph*, including living room, bedroom, kitchen, bathroom, and hallway. The objects in these zones, such as the TV set, lamp, toilet, oven, and bed, are mapped onto leaf nodes. Except for the electric current sensors, all sensors are grouped based on their location when extracting occupation of zones. For example, the PIR sensors for toilet and bathroom are grouped together, and the oven, trash, and kitchen are grouped together. The state changes of electric sensors are combined to those sensor events.

5.1.2. Settings of Data Reorganization

The threshold of time gap for event sequence splitting is set to be 2 minutes, which means if the time distance between two events is bigger than 2 minutes, then they are split into two episodes. The basic period is set to be *one day* (24-h), and the multi-period is set to be one *one week* (7-day). As mentioned in chapter 4, according to normal people's habits, the basic period time-line is split into six long segments, which are further split into 11 small segments. The long segments splitting is listed in Table 5.2. The time interval around the splitting point is defined as 2/3 length of the segment that is before the corresponding point. As shown in Figure 5.1, suppose there are two consecutive segment, Seg_i and Seg_{i+1}, and t_2 is the splitting point. The interval around t_2 consists of the last 1/3 of Seg_i and the first 1/3 of Seg_{i+1}.

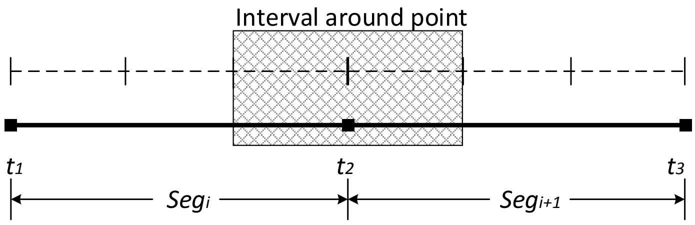


Figure 5.1.: Segment around splitting points.

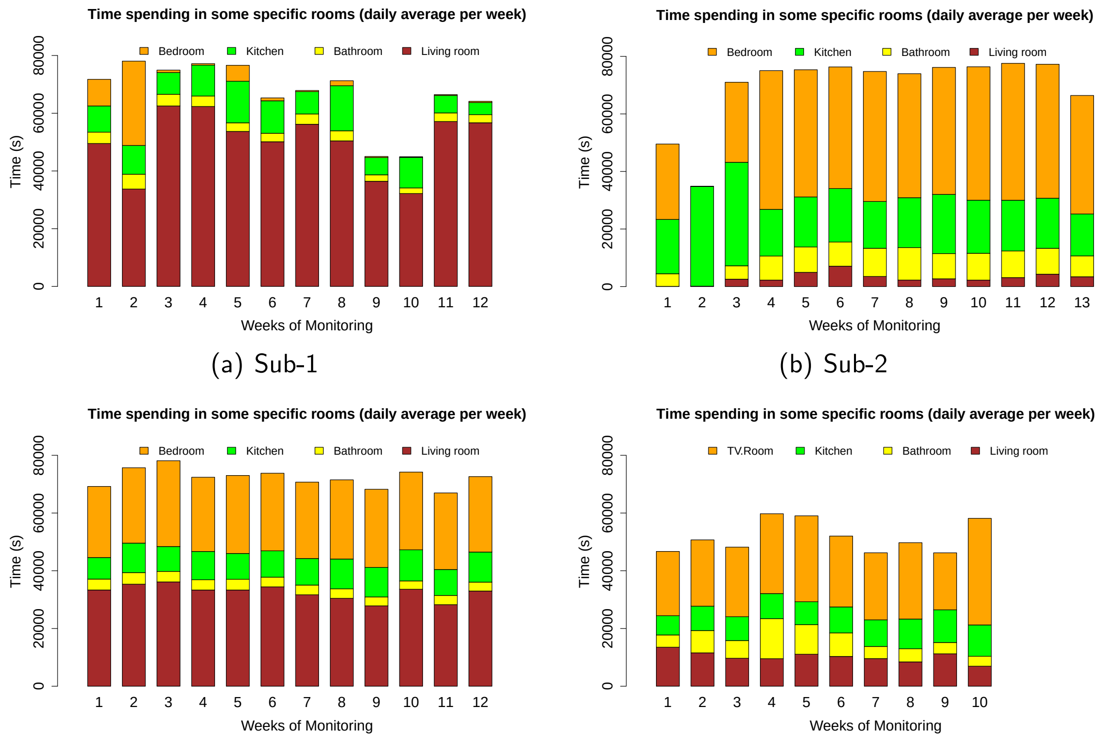
As for pattern discovery, the size of the *matrix mask* is adapted to the size of the segment. When the segment is bigger than 1 hour, there are 3 columns in the mask, otherwise there is only 1 column. The mask for the time interval around splitting points consists of 2 columns. After the event sequence splitting and packaging in *bags*, the *Apriori algorithm* is used to discover frequent event items, which are regarded as behavior patterns. In the R environment, the package *arules* implements this algorithm. The arguments are set as follows, target: "frequent itemsets", minimum support: 0.2; minimum length: 2; maximum length: 10.

5.2. Basic Profiling

After the processing of event extraction, some basic behavior information can be easily obtained through statistics. This section will present some results of subjects' basic profiling. Since the time-line is treated as a hierarchical structure, i.e., basic period and multi-period, the average statistic of a basic period over a multi-period is calculated to denote resident's behavior. Thus, in the current work, the statistic of one day over one week is presented.

5.2.1. Daily Average Time Usage Breakdown

Although we cannot recognize exactly what activities are performed through the sensor events, they can still provide useful information via zone usage since usage of specific rooms is closely related certain activities.



(a) Sub-1

(b) Sub-2

(c) Sub-3

(d) Sub-4

Figure 5.2.: Time spent in some specific rooms.

We say that some rooms' usage can represent resident's behavior, especially the B-ADLs, so as to provide information about a resident's health status. For instance, the frequency of toilet usage can manifest bowel and bladder control and management, personal hygiene related activities usually take place in the bathroom, and eating activities in the kitchen and dining room. In this section we calculate the mean value of amount of time spent in specific rooms of interest during one basic period, i.e., one day. Four rooms are chosen for each subject. Bedroom, kitchen, bathroom, and living room are chosen for subjects-1, 2, and 3. And a room where a TV set is located is chosen to substitute for the bedroom for subject-4.

The results of these four subjects' room usage are shown in Figure 5.2. From the results we can learn that: Firstly, variances in time usage ratios show up in all subjects. Subject-1 spent most time in the living room and what is unusual is there is little time spent in bedroom. Subject-2 hardly used the living room. Subject-3 spent comparable amount of time in both the living room and the bedroom, and subject-4 spent more time in the TV room. Secondly, as to each subject the ratio is generally stable, which also indicate that the residents are maintaining their own living pattern.

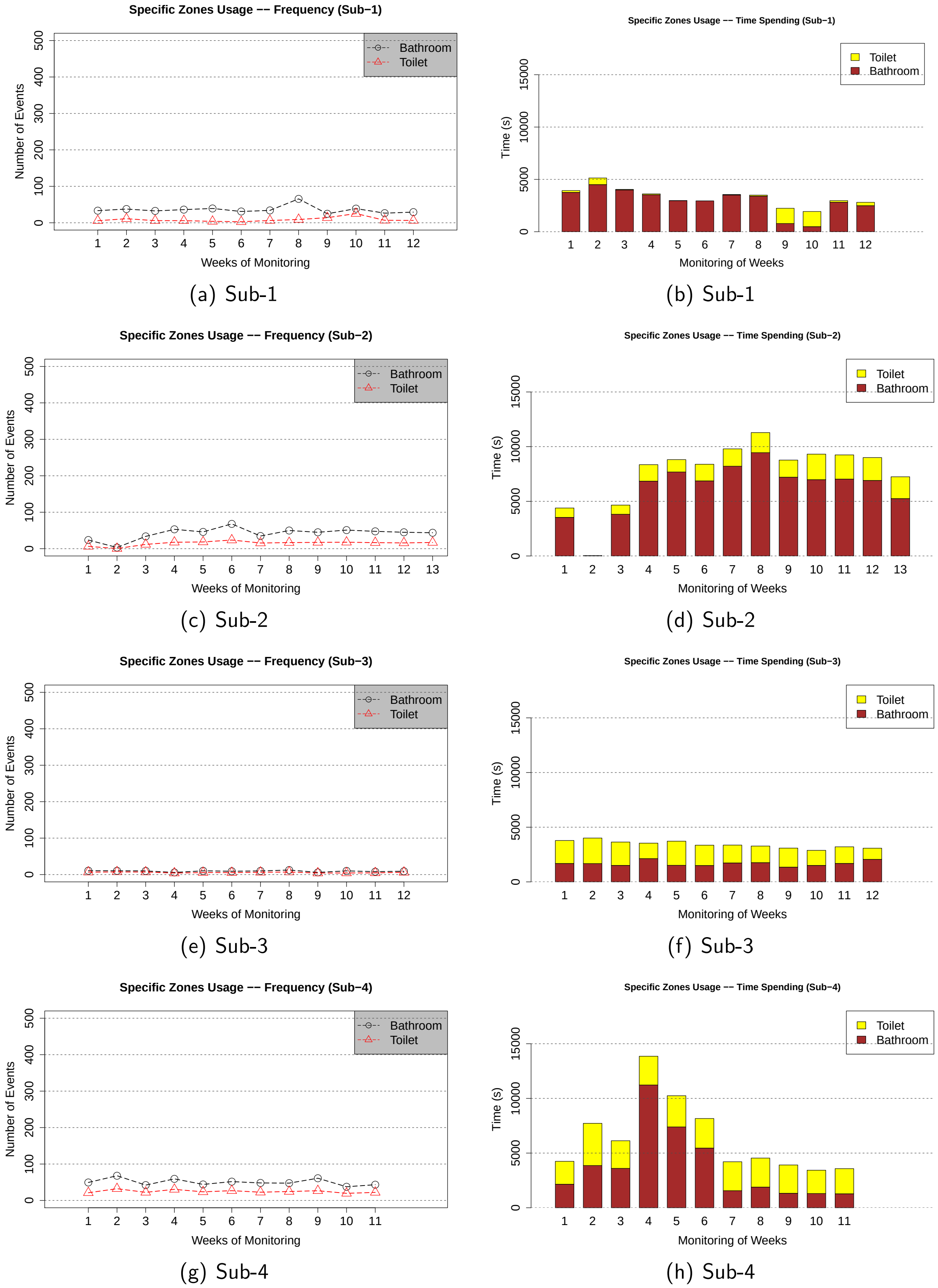
5.2.2. Usage of Specific Zones

Considering the subjects recruited for the GAL-NATARS study have been discharged from hospital following hip surgery, the mobility of these patients is obviously limited. The ability to accomplish self hygiene is an important means of assessing rehabilitation. Most self hygiene activities are carried out in bathroom, such as showering and washing of hands. Toilet usage is essential for people's daily life, which reflects the state of digestive system. Thus, it is expected that the situation of *bathroom* usage and *toilet* usage can provide valuable information on health status. In this section, the daily frequency of bathroom usage and toilet usage and the time spent in them over one week are investigated. The results are shown in Figure 5.3. The first column figures shows the frequency of bathroom and toilet usage of these four subjects, and the second column shows the total time spent in using each. Generally, they exhibit different behavior in both aspects, frequency and time spending.

To explore the relation of these two measurements, here we carry out the correlation analysis with Pearson's product moment correlation coefficient. As shown in Table 5.3, some subjects show higher correlation coefficients, with absolute ρ value close or bigger than 0.8, implying stronger correlation of corresponding item, either frequency or time spending. Subject-2 and subject-3 have higher correlation in both frequency and time spent, subject-4 only shows in *frequency*, and subject-1 doesn't show evidence of correlation in either *frequency* or *time spent* with the absolute ρ value smaller than 0.4. Through this we can conclude there is not a necessary connections between bathroom usage and toilet usage, even though they are in close proximity of each other. This mainly depends on residents' habits and condition. For example, if a resident is depressed due to anxiety about his or her body, he or she may be more restless and thus the bathroom might be visited more frequently.

Table 5.3.: The results of correlation analysis

Subject ID	Zone Usage Frequency		Zone Time Spending	
	$\rho_{X,Y}$	$p-value$	$\rho_{X,Y}$	$p-value$
Subject-1	0.098187	0.7614	-0.47518	0.1185
Subject-2	0.96788	< 0.05	0.84058	< 0.05
Subject-3	0.78058	< 0.05	0.81438	< 0.05
Subject-4	0.90558	< 0.05	0.39655	0.2273



(a) Sub-1 (b) Sub-1

(c) Sub-2 (d) Sub-2

(e) Sub-3 (f) Sub-3

(g) Sub-4 (h) Sub-4

Figure 5.3.: Specific zone usage.

5.2.3. Activity Level of Multi-period

As introduced in chapter 4, the events occurring on the central node in the *@home graph* are used to indicate the activity level. Based on the topologies of different *@home graphs*, the *hallway* node is usually selected as the central node. The average number of passages through the hallway per day over a week is calculated, which is simply the number of entering the hallway or exiting the hallway.

The results of average activity level of four datasets are shown in Figure 5.4. Subject-1 exhibits the highest activity level. Due to missing data, the second value of subject-2 is 0, and the first and third values are also lower. Aside from the outliers, subject-2 shows high activity level, which is similar to subject-1. Note that there is drop from week-6 to week-7. Subject-3 and subject-4 have much lower activity levels. Besides, it seems that subject-3 and subject-4 maintain stable activity level.

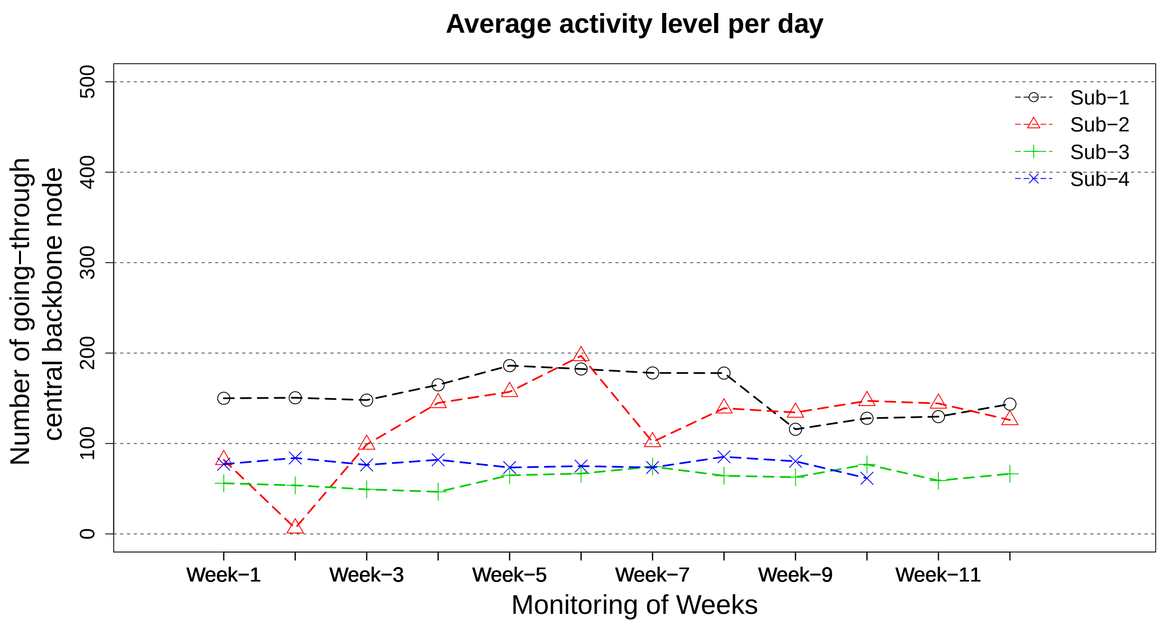


Figure 5.4.: Activity level represented by events triggering of the central backbone nodes.

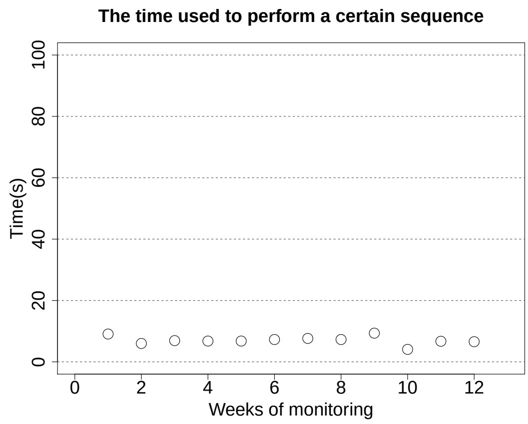
5.2.4. Results of Walking Speed Measuring

A specific sequence is chosen for every subject to assess the daily average walking speed. Based on the results of daily average total time spending (see section 5.2.1), we choose the sequence denoting long route connecting zones that are frequently used. As to subject-1, it is from the living room to the bathroom; as to subject-2, it is from the bedroom to the bathroom; as to subject-3, it is from the living room

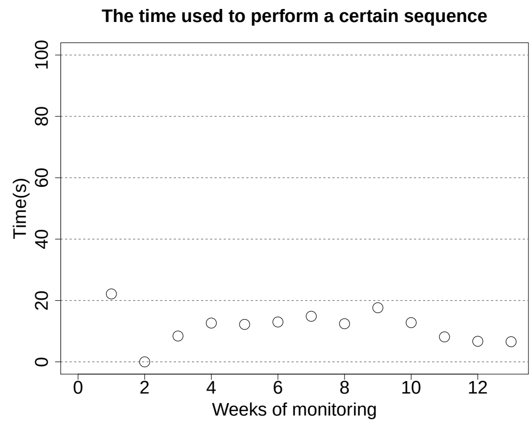
Table 5.4.: The specific sequence chosen for 4 subjects.

Subject	sequence
Subject-1	{ “LE”, “FS”, “FE”, “BS” }
Subject-2	{ “SE”, “FTS”, “FTE”, “BS” }
Subject-3	{ “LE”, “FS”, “FE”, “KS” }
Subject-4	{ “LE”, “FS”, “FE”, “BS” }

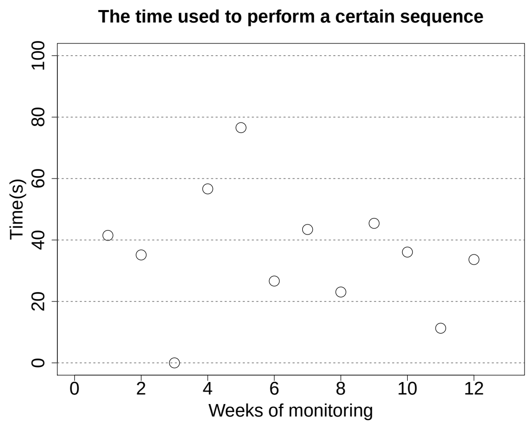
to the kitchen; as to subject-4, it is from the living room to the bathroom. They are listed in detail in Table 5.4.



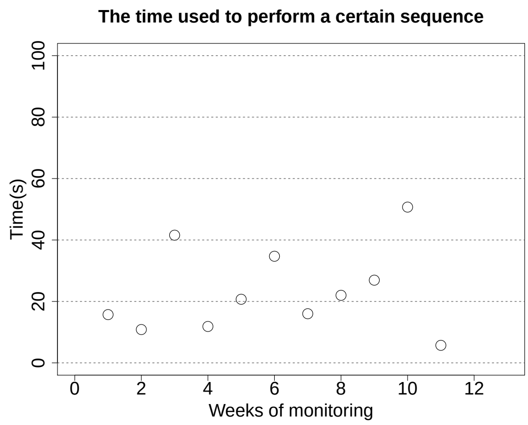
(a) Subject-1



(b) Subject-2



(c) Subject-3

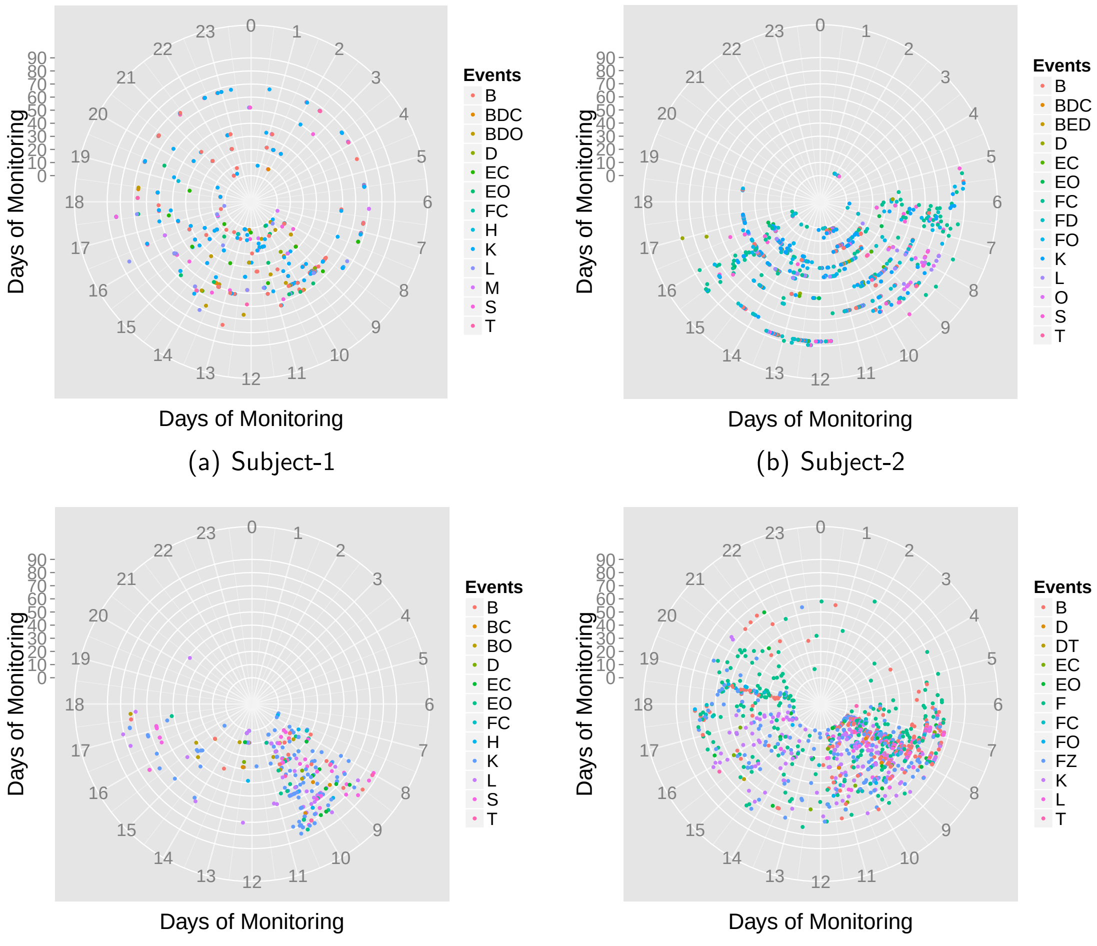


(d) Subject-4

Figure 5.5.: Time needed to accomplish a certain sequence.

The time needed to accomplish corresponding route is averaged in week and results are shown in Figure 5.5. Subject-1 and subject-2 represent smooth change of walking speed. As to subject-1 the values are around 10, and as to subject-2, the needed time is more in the first half monitoring term than the second half. However, subject-3 and subject-4 represent quite scattered distribution of time to go through the routes. This is not match the mobility related assessments results, which will be shown in section 5.4.

5.3. Multi-period Lifestyle Regularity



(a) Subject-1

(b) Subject-2

(c) Subject-3

(d) Subject-4

Figure 5.6.: Four subjects' Jump plots in spiral plots.

5.3.1. Results of Multiple Persons Detection

Before discovering subjects' lifestyle patterns, the datasets should be firstly investigated whether there are unexpected situations, and if there are, how these situations distribute. Here the *unexpected situations* refers to multiple persons showing up in the apartment or house. Based on the principle introduced in section 4.1.3, the jump points are marked. The results of these four subjects datasets are shown by means of spiral plot, which has been used in section 4.2 (see Figure 5.6).

The results illustrate that in general there are more *jump* points during daytime. There are less circles with *jump* for subject-1. With regard to subject-2 and subject-3, most of the *jumps* emerged during the daytime, and there is few points at night. Some regularities have been displayed w.r.t. subject-3 and subject-4. There are alway intensive *jumps* from 08:00 to 10:30 for subject-3, and there are always more intensive *jump* points around 08:00 for subject-4. The results might be explained by the fact that some caregivers coming periodically (e.g., daily). As for subject-2 and subject-4, considerable number of *jumps* occur at night, even the density is much lower than daytime. This is not reasonable, which may be due to the sensitivity of used sensors.

5.3.2. Representing Lifestyle

As there is variance in living environments and residents, there may be different sensors used and the sensors may also be deployed in different positions. Even for those residents with similar health conditions such as the subjects recruited for the GAL-NATARS study, there is still some variation in sensor deployment. After event extraction and data reorganization, pertaining to a certain resident, a set of interesting event sets must be defined before behavior patterns can be deduced from the sensor data. In the current work, events connecting to basic self-care activities are selected to combine the interesting event sets (IES).

For subject-1, the interesting event sets are selected and listed in detail, followed by their intentions. The abbreviations are defined using the following rule: the first letter corresponds to a specific part of the words of living environment. For example, ‘L’ is for *Living room*, ‘K’ is for *Kitchen* and so on; the last letter ‘S’ and ‘E’ stand for *Starting* and *Ending*, respectively. The meanings of the abbreviations are interpreted in Table 5.5. Similar work was also carried out for each of the other subjects.

- { “LE”, “BS”, “TS”}: to reflect the hygiene, bowel and bladder control and management;
- {“LE”, “KS”}: to reflect completion of house task;
- {“LE”, “BE”, “KS”}: to reflect eating and hygiene;
- {“HS”, “KS”}: to reflect eating;
- { “LE” “HS” “KS” }: to reflect eating;

- { "SE" "BS" "TS" }: to reflect bowel and bladder control and sleeping.

Table 5.5.: Interpreting abbreviations

Abbreviations	Meaning
BE	Bathroom Ending
LE	Living room Ending
KS	Kitchen Starting
HS	Stove (German: Herd) Staring
TS	Toilet Starting

The IESs are used as the target input for the pattern discovery procedure (see section 4.2), which is based on an association rules discovery algorithm. When implementing this method, the following processing is conducted: if the support value of an event set is lower than the threshold in a certain segment, minimum support, which is 0.2 in this experiment, and a fixed value, 0.1, is assigned to those cases below minimum support. To illustrate the subject's behavior patterns, a part of the results from subject-1 are represented in the form of a bar plot (as shown in Figure 5.7). For the full results of each subject's pattern, please refer to appendix A. Note that the threshold is shown by a red line. For the first impression, we can see that: 1). the subject uses the bathroom and toilet late at night (from 00:00 to 03:00); 2). for this subject entering the living room and kitchen also occur late at night (from 00:00 to 06:00); 3). cooking usually happens in the daytime from 08:00 to 18:00; 4). The patterns are not very stable and there is some change across the observation term.

As introduced in chapter 4, in a multi-period, some segments are usually combined to form an interested time interval (*iti*). This work will formulate three *iti*s, *morning*, *evening & night*, and *late night*, where *late night* combines *late night 0* and *late night 1*. This lies in the following three points. Firstly, more activities are likely to be intensively conducted in the morning after getting up, and during the night before going to bed; secondly, the activity during late night can represent the subjects' sleeping quality, which is a very important indicator of the subject's rehabilitation; thirdly, there is likely to be less disturbance due to other visitors in these intervals, hence the events occurring in these intervals are more likely to represent subjects' behavior pattern.

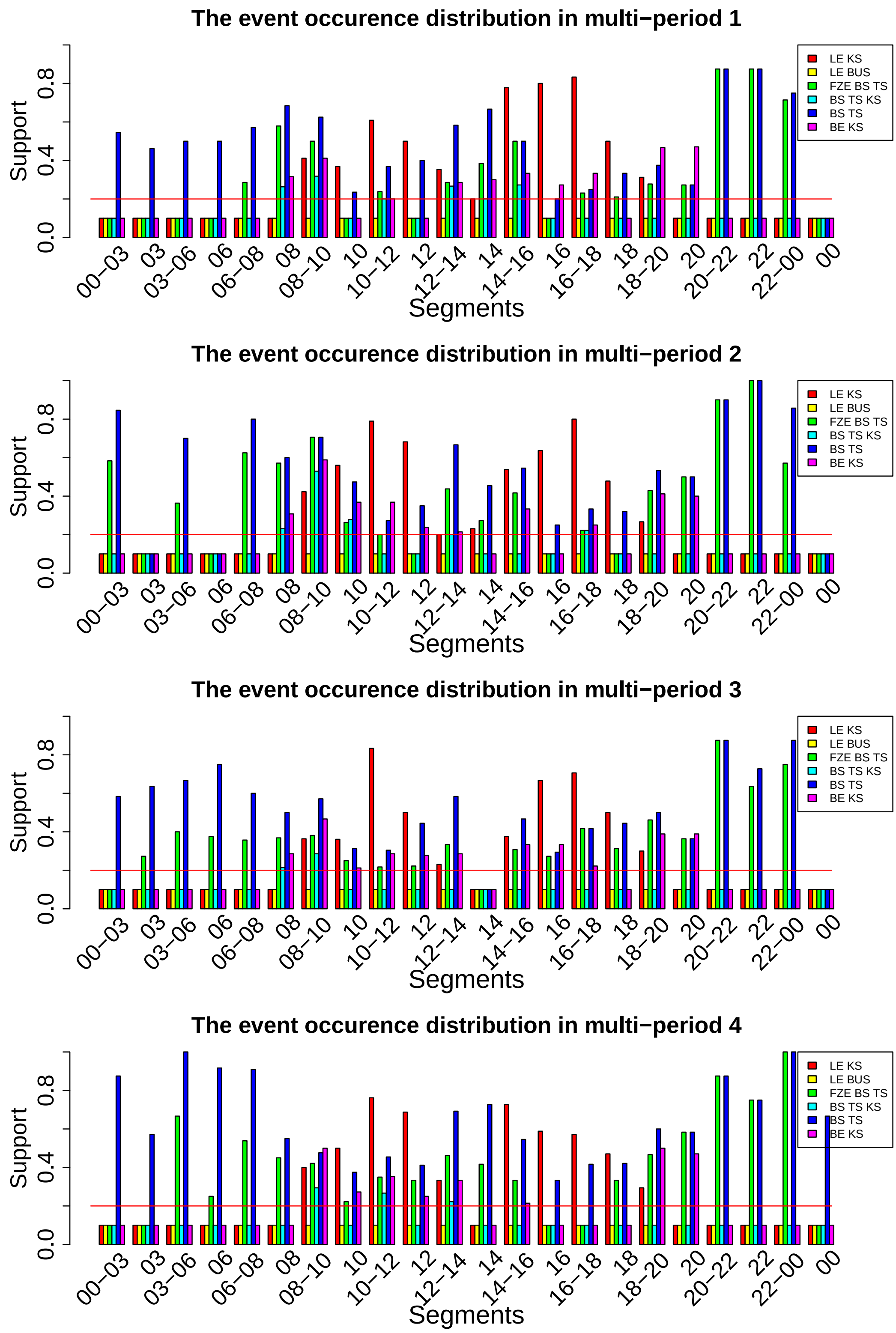


Figure 5.7.: **Patterns of a number of consecutive multi-periods of subject-4.**

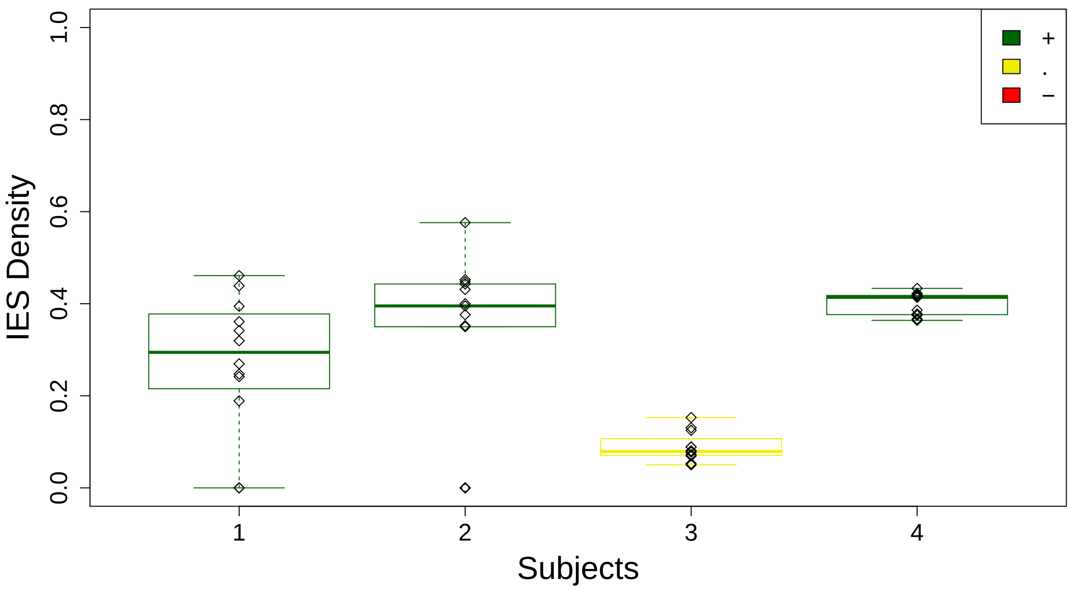
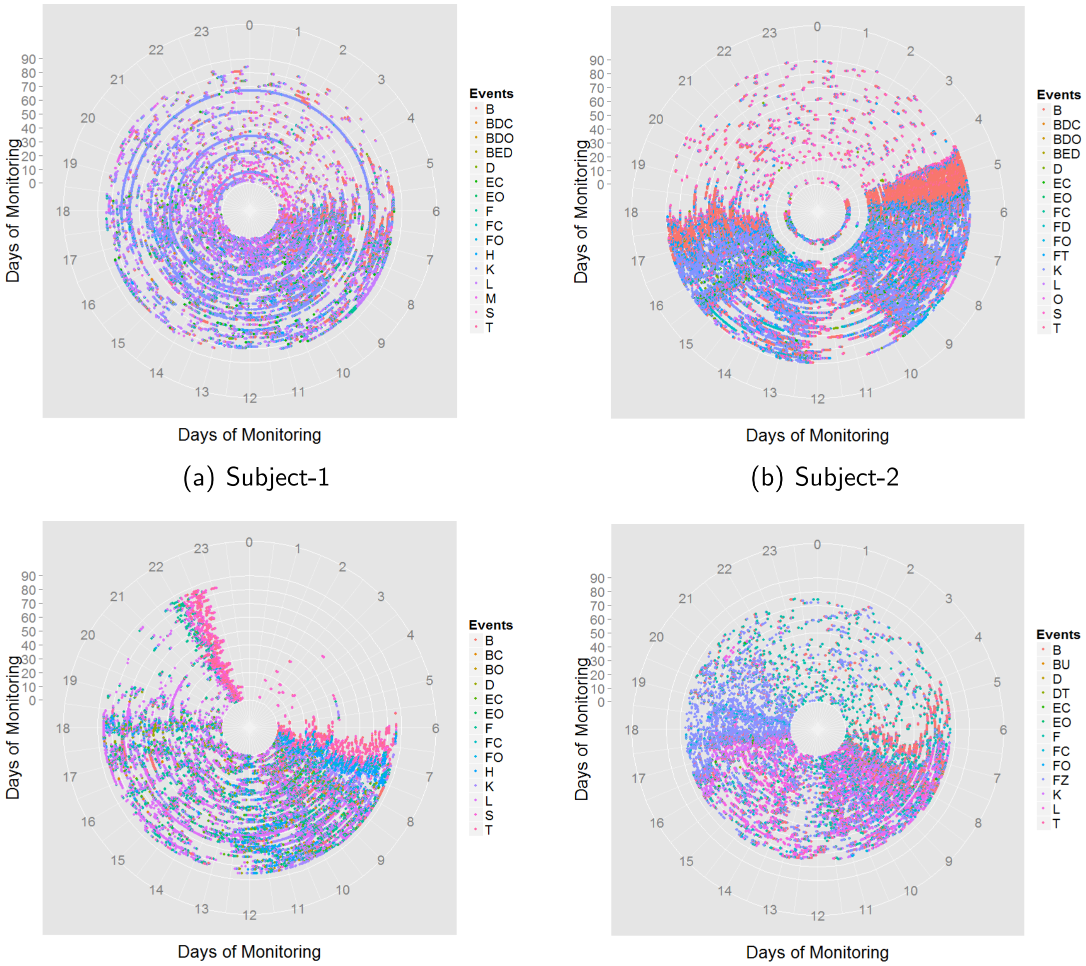


Figure 5.8.: IES density over interested time intervals.

As long as the evident IESs in one *iti* are discovered, they are used as attributes to characterize behavior in corresponding time interval. According to the definition, IES density over the monitoring term is calculated. For every multi-period (i.e., week), the IES densities of all *iti*s is calculated and the results are combined and normalized. To illustrate the difference of these four subjects' IES density, a box plot is used to present the calculated results as shown in Figure 5.8. It illustrates varying distributions, which can be categorized primarily into two groups. Subjects-1, 2, and 4 show higher IES density, while comparatively fewer IES are performed by subject-3. The box color illustrates the subjects' health status, which are evaluated by nurses. Green color indicate improved health, yellow indicates stable health, and red indicates decreased health. The correlation between IES distribution and subjects' health status will be discussed in section 5.4.

5.3.3. Change in Lifestyle

To obtain an intuitive sense of the subjects' activity during the observation term, the raw sensor records of all subjects are firstly visualized by means of spiral plot (see Figure 5.9). As with the examples in section 4.2 and 5.3.1, the clockwise circle illustrates the time of each day, from 00:00 to 00:00 of next day. And each circle shows



(a) Subject-1 (b) Subject-2

(c) Subject-3 (d) Subject-4

Figure 5.9.: Four subjects' sensor data visualized by means of a spiral plot.

one day's sensor records. From Figure 5.9, it is evident that each of the subjects' spiral plots have different appearance.

Subject-1

Subject-1 has an irregular lifestyle and the spiral plot appears more irregular. There is no evident separation between daytime & nighttime, and the density of sensor triggering is similar between daytime and nighttime.

Subject-2

The spiral plot of subject-2 shows very different appearances between daytime and nighttime. It seems this subject has more regular lifestyle. There is a high density of sensor triggering during daytime, while the density is quite low and consistent during nighttime. Also, some sensors are intensively triggered in the same time interval, around 05:00 and from 17:00 to 18:00.

Subject-3

Subject-3's spiral plot also shows different appearances between daytime and night time. However, compared with subject-2, there is nearly no sensor triggering activity at night. There is also intensive sensor triggering at similar time intervals, from 06:00 to 07:00, and from 22:00 to 23:00.

Subject-4

The spiral plot of subject-4 also shows distinctions in sensor triggering between daytime and nighttime, and the border is occur around 06:00 and 17:30. Moreover, there are relatively more sensors triggered by subject-4 than those triggered by all other subjects except for subject-1.

According to the definition of *change* in behavior patterns, we calculate the change w.r.t. every multi-period (week). The *change* is also calculated in the same *iti*s as the calculation of IES density, i.e., *morning*, *evening & night*, and *late night*, and the results are summed up for each multi-period. Thus we obtain a single value for every multi-period. For clarity, all results are normalized on a scale from 0 to 10. All subjects' change is presented in Figure 5.10. The four subjects are denoted in different symbols and colors. The scale marks of horizontal axis denote the beginning of multi-periods (i.e., week). For example, the point 2 means the starting of the second week. The change between two weeks, say $Week_{(i)}$ and $Week_{(i+1)}$, is plotted at the beginning of the second one, i.e., $Week_{(i+1)}$. The change of subject-1 has significantly declined in the first five weeks. It has to be declared that during week 9, 10, and 11 the subject wasn't at home because of Christmas. The results show unusual change due to sudden data disappearing and arising. From the raw data visualization it is evident that there is data missing in the first weeks of subject-2, which leads to no patterns being discovered in these three weeks. When there are patterns evident in $Week_4$, high change come out between $Week_3$ and $Week_4$. Because of data missing, there is no change in the first two weeks, and there also huge change when there is data measured. The first three values should also be excluded when considering the results with health. Except the first three points, in general subject-2 has shown decreasing trend of change, which means his/her living was getting more regular. Subject-3 shew deceasing trend, but after that the change gradually increases. Subject-4 performed the best of keeping regular living. The change is alway under 2, even though there is slightly increase in the last couple weeks.

In order to analyze the difference in change of the subjects, the results are also plotted by means of a box plot (see Figure 5.11). By looking into the change shown in Figure 5.10 and Figure 5.11, the results are summarized. Generally, the change of subject-1 is the highest, except one outer. This is also in line with the raw data visualization in Figure 5.9. Subject-2 and subject-3 exhibit similar mean value of *change*, which is lower than subject-1, except the change between $Week_3$ and $Week_4$ of subject-2. Obviously, subject-4 shows the least change. Besides, not only are the values low for subject-4, but also the distribution is in a narrow interval, as shown in Figure 5.11.

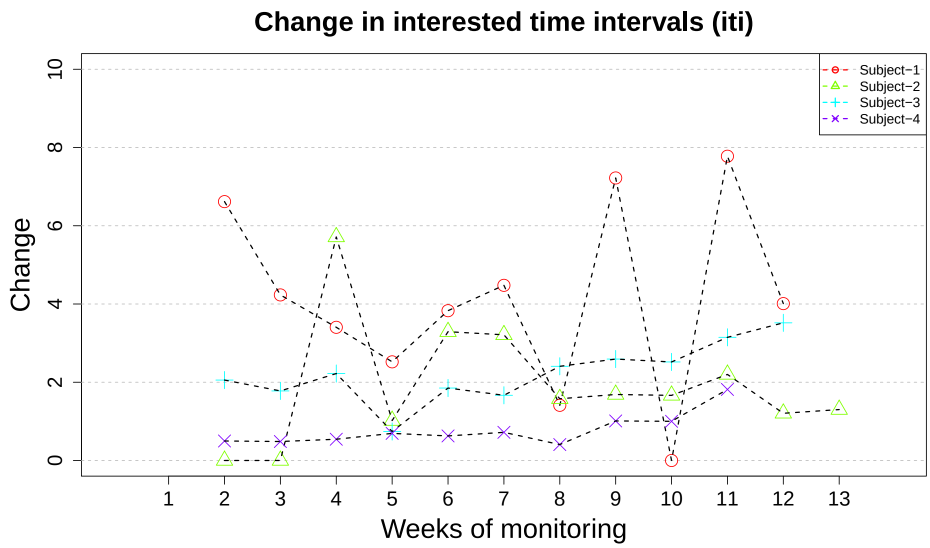


Figure 5.10.: Change in the observation term of all subjects.

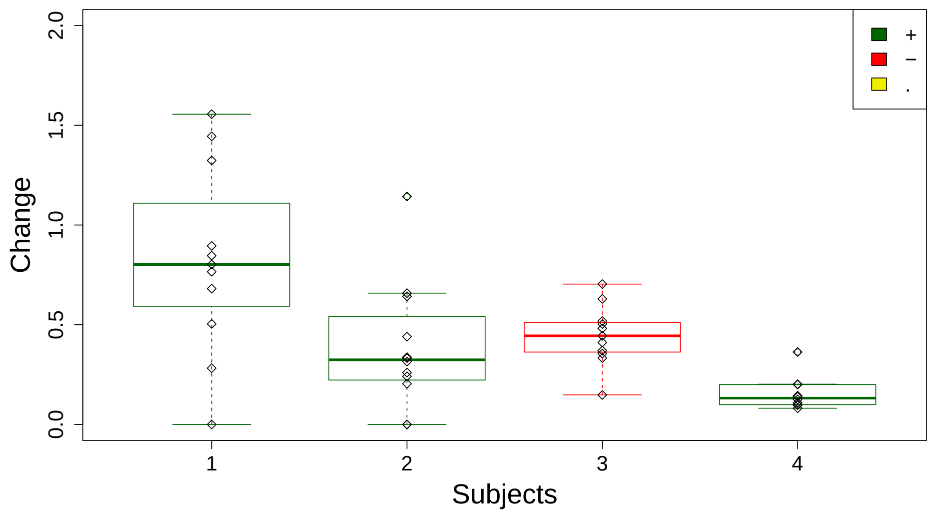


Figure 5.11.: All subjects' change in behavior patterns during interested time intervals, which are morning, evening & night, and late night.

5.4. Correlation with Medical Assessment Data

For the sake of connecting the information discovered to the subjects' health conditions, and evaluating the effectiveness of the methods proposed, a total of five kinds of assessments are taken into account, including Timed Up and Go (TUG), Short Physical Performance Battery (SPPB), Tinetti I and II, Barthel Index (BI), and Visual Analogue Scale (VAS).

According to the study plan of the GAL-NATARS study, medical assessment tests are performed a sum of four times. Data corresponding to four subjects is analyzed. Duration of monitoring term is three months, which is relatively short for generally modeling behavior change. Because of these limitations, a descriptive case analysis is carried out, hoping to provide a paradigm for inspecting a patient's health status from the behavior point of view.

Table 5.6.: The subjects' health state

Subjects ID	State
Subject-1	+
Subject-2	+
Subject-3	•
Subject-4	+

+: improved; •: stable; −: decreased.

First of all, according to the nurses' points of view and the assessments results, the subjects' health states are marked, and they are listed in Table 5.6. Four of the assessments are mobility related, including TUG, SPPB, Tinetti I and II, and BI. VAS is used to assess the subjects' feelings of pain, and the higher VAS value implies the subject feels more pain. Since these assessments have different scales, it is necessary to normalize them to a uniform scale for convenient processing. In the current work, they are normalized on a scale from 0 to 10, which is the same as the scale of change. For the original results of the assessments of the four subjects, please refer to appendix B. Furthermore, those four mobility related assessments are integrated into one value, $m.val$, to represent the subject's mobility. For initial processing, same weight is assigned to each of the assessments (see Formula 5.1).

$$m.val = \frac{1}{4}TUG + \frac{1}{4}SPPB + \frac{1}{4}Tinetti + \frac{1}{4}BI \tag{5.1}$$

The higher the $m.val$, the better performance of the subject's mobility. Taking the information discovered in this work into account, the descriptive analysis will be conducted from the following aspects:

- Activity level
- Evident IES density
- Change in behavior pattern

5.4.1. Activity Level

The activity level is obtained by calculating the number of passages through the central node on the *@home graph*, where the hallway is usually selected as the central node. For subjects with disease on the hip, the activity of passing through the central node reflects their ability to move around in the living environment, thus indicating the subject's rehabilitation. By comparing the marked subjects' health states with the activity level (Figure 5.4), the correlation is summarized into two groups:

- The subjects with *improved* state (subject-1, 2, and 4) has a higher activity level and the variance is more significant;
- The subject with *stable* state (subject-3) has lower activity level, and the trend remains stable or slightly declines.

5.4.2. Evident IES Density

From the analysis of activity level, those subjects with *positive* health status have relatively high activity levels. Intuitively, a higher activity level leads to a greater possibility of representing more regular behavior patterns. From the results of Evident IES distribution shown in Figure 5.8, we can also easily find the coincident situation. The subjects with improved health status also show higher IES density in the selected three segments, while the ones with stagnation and deterioration don't show as much.

5.4.3. Change in Behavior Patterns

Based on the processing of assessment test results, there are two values to indicate the subjects' health condition, i.e., the *m.val* indicating the mobility, and VAS indicating the feeling of pain. To intuitively show the correlation between the assessment results and the change value obtained through the methods proposed in this thesis, they are presented together in the form of a single chart with three curves (see Figure 5.12). In the subsequent paragraphs, the results of each subject will be described separately.

Subject-1

Outliers of change: at weeks 9, 10, and 11. Due to the subject's absence during those weeks, the patterns and change at those points are not reliable.
In the first five weeks, the behavior pattern change decreased from 6.61 to 2.52. Meanwhile the mobility increased by around 1 unit. In the following four weeks, the change increased, and a shallow valley was shown at the third assessment test. Note that the VAS generally showed an increasing trend, especially during

the last five weeks when the feeling of pain is much stronger than the weeks before.

Subject-2

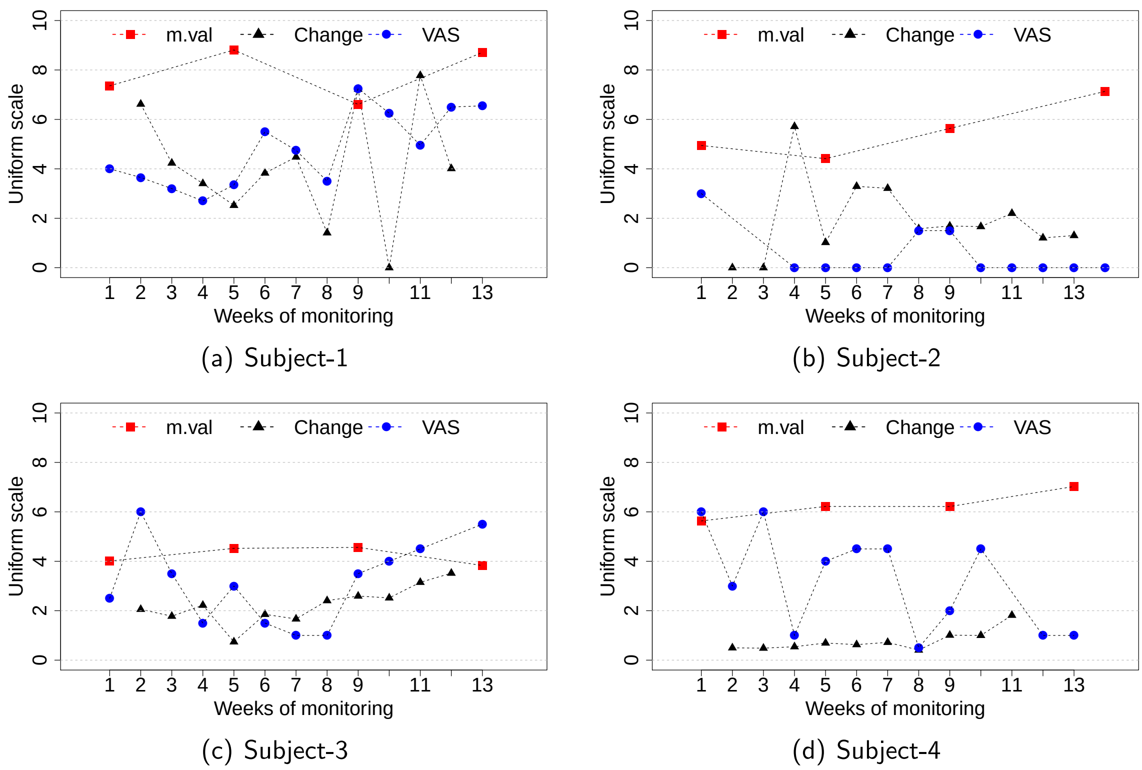
Outliers of change: at 2, 3, 4. Due to some unknown reasons, there is data missing in the first 3 weeks. The first 3 change values are not reliable.
Apart from outliers, the change generally exhibited a decreasing trend. Meanwhile the mobility represents gradually a increasing trend. Note that this subject has shown very low VAS compared with the others.

Subject-3

This subject's mobility is relatively low compared with the other three. There is nearly no improvement in mobility, with a slight decrease evident in the final weeks. The subject's behavior change showed increasing trend in the last few weeks. The VAS shows that the subject also felt significantly increasing pain in the last few weeks.

Subject-4

Through the whole observation term the subject exhibited very little change, which is no more than 2, and the *m.val* maintained in a high level and increased. The VAS fluctuated a lot during the monitoring period.



(a) Subject-1 (b) Subject-2 (c) Subject-3 (d) Subject-4

Figure 5.12.: The correlation between change behavior patterns and medical assessments.

By connecting the three results, some conclusions can be drawn. Less significant change usually accompanies better mobility performance, such as during the first five weeks of subject-1, subject-2, and subject-4. However, the change of behavior pattern is not the dominant factor to represent the subject's rehabilitation. For subject-1, the change in the last half term is high, but he or she can still perform well mobility. This mainly depends on the individual's health condition. On the one side, subject-1 showed high mobility, which means he or she was able to move and conduct more activity; on the other side, he or she expressed feeling more pain, which may have contributed to his or her irregular and restless lifestyle. In another case, subject-4 showed the least change, but there are some higher VAS values. Meanwhile, the *m.val* is not so high as subject-1. This suggests that the feeling of pain did not bother him or her too much, and this subject can maintain a stable lifestyle and make some improvement physically. Considering these three aspects, *m.val* represents the subject's ability to carry out some activities, and VAS probably has more influence on subject's psychology. Both mobility and psychology are able to influence behavior patterns, the variance of which is expressed by *change* in this thesis.

5.5. Chapter Summary

For the purpose of evaluating the methods and theories, the methods and theories are applied to the datasets collected from the GAL-NATARS study. This chapter firstly states the settings of experiment in 5.1. Some basic behavior related information is presented in section 5.2 and the lifestyle regularity is discovered and shown in section 5.3. To investigate the results, a descriptive analysis is carried out in section 5.4.

Chapter 6

Discussion

6.1. Achievements of the Objectives

In the context of ambient assisted living, this thesis has proposed a paradigm to extract behavior related information using unsupervised methods from datasets obtained through sensor-enhanced living environments. The theories and methods proposed in this thesis have been applied to the datasets of GAL-NATARS study, where the data is collected in real-life situations. Two types of information are extracted, consisting of basic profiling and behavior patterns. Basic profiling refers to some metrics on the basis of sensor events, such as time spent, usage of zones, activity level, and walking speed. They are obtained mainly through statistic calculation. Behavior patterns are extracted through discovering frequent event sets, and these patterns are regarded as attributes of a certain period. Through defining the change of patterns over time, the resident's ability to maintain the lifestyle is quantified as well. By analyzing the correlation between the information discovered in the current work and the results of assessment tests, this thesis has shown that the methods are able to provide complementary information to the health assessment, thus contributing to fully investigating the resident's health status.

During the whole procedure of data processing, only unsupervised methods are used. Labeling data is not necessary any more. The current work is on the basis of the assumption that the triggering of sensor is closely related to certain activities. The indirect approach is actually implemented to assess the resident's behavior. The primary information is identified through sensor ID, which shows which event is detected. To acquire the location of a certain sensor, the layout of a apartment or house is in demand. Training model can also be avoided. Each of the subjects are regarded indistinctly excepts for specific interesting event sets. For any subject, the data is input to the algorithm after preprocessing. Additionally, users benefit from these advantages. The resident can maintain the living as normal and there is no additional task, such as recording any activities. There is no need to carry out any training work before the methods are used for a different resident, which saves much time and reduce the complexity of study. The accuracy is equal for different residents since there is no special parameters corresponding different residents. The sensors are directly deployed in resident's living environment and the data can be analyzed uniformly.

6.2. Answers to Questions

To reach the objective, a number of questions have been defined at the beginning of this thesis (see section 1.4). This section will discuss the current work from the aspects of these questions, and to see how far they have been addressed.

Q1. How can living environments be formally described?

The architecture of living environment, including objects inside, is mapped onto a graph named as *@home graph*, by which the topology can be intuitively illustrated. The triggering of sensors is regarded as the state change of components, which may be either nodes or edges of the graph. In this way, some unnecessary information is ignored and algorithm settings can be easily configured.

When the living environment is mapped onto the *@home graph*, the size of zones and objects and the distance between them are not taken into account. The issue of interest is the relation of location. Therefore, some quantified measurements related to size and distance are beyond the ability of this kind of abstraction. For instance, the exact speed of passing through a certain zone cannot be calculated.

The interactions between resident and living environment are abstracted to two kinds of events, i.e., time-point-based events and time-interval based events. Based on the time attributes of events, we can easily get the information about the resident's basic behavior, such as activity level, time spent in specific rooms, and usage of specific zones. The abstraction neglects concrete sensors, and this is able to orient to normalized data processing. This general abstraction can also be used in any other related topics.

Q2. What kind of data preprocessing must be conducted?

Aiming to deal with the issue of multiple data formats reported through different sensors, data normalization has been implemented in this work. By normalizing all kinds of data into a sequence of events, it is convenient to process them through uniform approaches. To prepare input data, the dataset is reorganized through two steps, consisting of time-line splitting and event sequence clustering.

Firstly, it is reasonable to split the event sequence into episodes. Since people perform activities in a kind of rhythm, comparable shorter distance usually exists between some events. Intuitively, the events in the same episode have more correlation. Clustering those events is able to highlight their relevance. The events in one episode potentially imply some specific activities, especially some basic self-care activities, which are of most interest in this research.

Secondly, it is reasonable to split the time-line of one day into some segments based on normal people's everyday routine. By splitting the time-line into segments, sensor events can be considered with time intervals. For instance, if the sleep quality is of interest, the events within intervals at night should be analyzed; if the eating activity is the concerned issue, the events in the kitchen during daytime should be paid more attention. The time-line splitting can be improved as well. In the current work, it is carried out on the basis of normal people's lifestyle. However, the resident, especially the older people or the ones with health problems, probably have very different habits in real-life situations. From the spiral plot of Subject-1's raw sensor data, for instance, it is evident that there are considerable events at late night; and from the IES_{evi} distribution discovered via the current work, as for subject-1, a number of events turn out to be regular at late night, which shows this person is restless. For subject-4, from the results of IES_{evi} distribution, it shows that the *Kitchen* related events are likely to emerge in intervals 10:00 – 12:00 and 14:00 – 18:00, which is slightly different from the normal people's eating time. It would be helpful to implement adaptive time splitting. In order to more precisely show the distribution, the size of time interval can be smaller when there are more events of interest, otherwise, bigger interval is enough for the situation with fewer events occurrence.

Thirdly, it is reasonable to use hierarchical periods. It is useful to reflect the resident's lifestyle regularity. If the resident is following a regular life, there is higher possibility for him or her to perform a certain activities more frequently in the same time intervals (i.e., segments) across some periods (i.e., multi-period). Regularity is a statistics term, and it is necessary to include the information of some basic periods so as to assess the regularity during a certain duration. So in the current work the whole observation term is split into some multi-periods. On the one hand, the regularity performance can be assessed in each multi-period through the distribution of evident

IES. On the other hand, the situation of change can also be assessed across multi-periods. Only frequently occurring events are shown, which is helpful to reduce the influence of incorrect data. It is difficult for the resident to maintain a regular living for a long term, especially for those with health problems, such as mind impairment and lack of physical ability. So the ability to maintain regular and healthy lifestyle is seen as an important indicator to assess the patient's rehabilitation.

Q3. How behavior patterns can be extracted from datasets?

This thesis mainly makes use of *apriori* algorithm to discover frequent event sets. To obtain wished results, an input data construction scheme has been designed (see section 4.2.2). Based on time-line splitting, in each segment, all possible combinations of subsegments are traversed, and the combinations with support value over threshold are regarded as evident. Through this scheme, first of all, more precise distribution of the evident IES can be discovered; and the influence of uninteresting events can be reduced. However, the time consumption is high w.r.t. this approach. As introduced in section 4.2.2, for a segment with size $s \times n$, the number of iteration is s^n. As the segment is split into more subsegments, much more time will be needed to traverse all possible combinations.

The *jump* has been carried out to investigate the situation of multiple persons (see section 5.6). From the results it is obvious that some subjects exhibit considerable *jump* points occurring at night, such as subject-4. As known, the inclusion criteria state that only the patients living alone are included, which means there shouldn't be so many *jump*s during nighttime. The reason of the results may be also due to the sensor sensitivity. Because of missing of triggering, the algorithm detects more unexpected points. This implicates in the future study, achievable and practical approaches to deal with multiple persons, such as the situation of visiting, should be designed.

IES (interesting event sets) for each of the subjects have been defined. The evident IES, which is a subset of IES, in the chosen time interval is adopted as the pattern of corresponding interval. Event sets need to be refined w.r.t. specific users, which is critical to optimally manifest the resident's health status. At present, they are determined based on the medical assessment items that are closely related to physical ability and behavior. However, IES can be refined by consulting with clinicians. Moreover, by using the scheme of information discovery proposed in this thesis, the statements proposed in chapter 4 can be filled out. Suppose the sensors are fixed in position, the sensor ID can identify which sensors are triggered and where these sensors are, which provide the information of *where* the resident is and *what* they are possibly doing. The only problem is to answer *when* some certain activities occur. By time interval analysis, the most frequent events during a multi-period are shown, resulting in what kinds of events are usually done during which time intervals. Thus, the statement, "the resident usually does ... (what) ... (where) ... (when)", are filled out. Through the whole procedure, only unsupervised methods are used to extract behavior related information, which has been discussed in section 6.1. Neither data labeling nor model training is needed.

However, there is still some room that can be improved. Since we deal with the time-line in a hierarchical way, i.e., basic period and multi-period, it is assumed that the resident's behaviors are periodically base on basic periods. One limitation of the way to reconstruct the time-line cannot discover those events whose period is bigger than basic period, such as the same as multi-period. In this situation, those events will be regarded as rare, and they cannot be shown. For instance, the laundry is also an important indicator to resident's self care, and one way to detect it is to monitor the usage of wash machine. However, as for those persons living alone, the wash machine is usually used weekly whose period is much bigger than basic period set in this work, *day*. They cannot be discovered merely through the methods of this thesis.

Q4. What kind of metrics can be selected and defined?

In the current work, four metrics are defined (see section 4.3). From the results shown in chapter 5, it is evident that they have shown variant performances.

Through statistic methods, basic profiling depicts the resident's general behavior information. After data preprocessing, the raw data is converted into an event sequence that is able to more intuitively represent resident's behavior. Based on the event sequence, it is very easy to calculate some metrics.

Daily average time usage breakdown shows the proportion of time spent in different rooms and how the proportion changes over time. This metric is able to represent the resident's habits to some degree. For instance, subject-1 spent most time in living room, but quite little time in the bedroom (see Figure 5.2 (a)). This result might be explained by the fact that this subject preferred to rest in the bedroom rather than bedroom. And in another case, oppositely, subject-2 hardly used living room (see Figure 5.2 (b)). This may be explained by the fact that the subject didn't have so high mobility as subject-1, and he or she preferred to stay in bed. Time spent also implicates that the residents can generally maintain their own living patterns. The obvious difference between these two subjects also confirms the statement, "Different individuals can exhibit different impression through data distribution", which is proposed in section 4.2.

The usage of bathroom and toilet (specific zones usage) is expected to be able to indicate the resident's bowel and bladder control and management. From the analysis of correlation between frequency and time spent in bathroom and toilet usage (see Table 5.3), we concludes that some subjects tend to use toilet when he or she enters the bathroom, such as subjects 2, 3, and 4, while the others exhibits less relevant connection. The results suggests that the usage of toilet and the usage of bathroom should be considered separately. It is possible for a restless resident just to go around in the apartment, but meaningful activities.

The resident's walking speed has been assessed in the context of GAL-NATARS study. However, the results turn out to be unreliable. Some residents show quite scattered plots, which is not in line with the assessment results. There are two most possible reasons: 1). interference from visitors. 2). the lower accuracy of sensors. For reason 2, the motion sensor (PIR sensor) used in the GAL-NATARS study is not a very precise sensor, and its triggering is influenced a lot by some causes, such as the sensor deployment and the direction of movements. If more accurate detection is wished, other sensors that are able to detect the signal under strict conditions should be chosen, such as light barrier.

In the current work, evident IES w.r.t. chosen time intervals are regarded as attributes to represent the resident's behavior patterns, as demonstrated by equation 4.1. Considering the IES density, the *change* is defined based on Hamming distance to assess the variance of behavior patterns over time (see equation 4.8). Given two sets with the same Hamming distance, the change is higher w.r.t. low IES density. The change has qualified the variance of pattern in a scale from 0 to 10, which is able to clearly show the range of variance. From the exhibitions of all subjects, there is obvious difference in ability to maintain their typical behavior. The correlation between change in behavior patterns and the subjects' health status will be discussed in next paragraph (Q.5).

Q5. How can the relation between the behavior pattern obtained through the proposed methods and the results obtained through medical assessments be evaluated? Furthermore, how can the effectiveness of the proposed methods be assessed?

According to the nurse's opinion, each of four subjects are marked as either "improved" or "stable". Towards convenient comparing, both assessment results and algorithm results have been normalized on the same scale, from 0 to 10. The correlation has been analyzed by comparing the information discovered through the proposed unsupervised methods and the results obtained through medical assessments (see Figure 5.12). For subjects with better recovery, there is higher activity level and IES density, and the change also tends to be less.

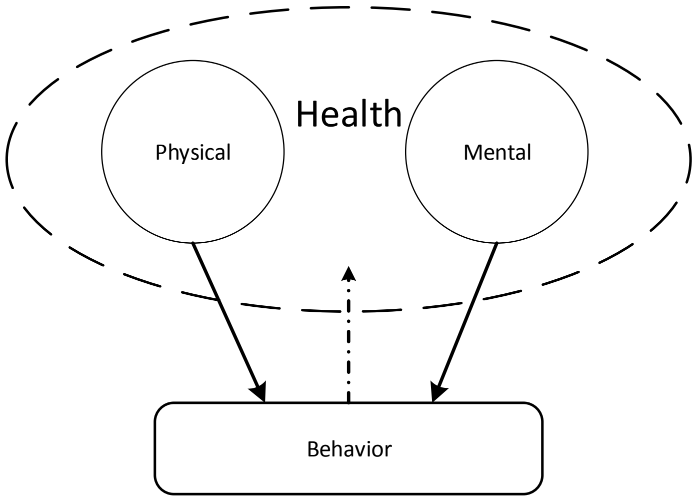


Figure 6.1.: **Relationship between three aspects.**

Through the analysis it has been shown that the discovered information is able to supplement and contribute to the diagnosis, but determinant relation between behavior and rehabilitation has not shown in the results. Human health status can be regarded as consisting of at least two aspects, i.e., physical health and mental health. Physical health determines the ability to conduct some movements, i.e., the mobility, while mental health determines the way to perform activities. Human behavior depends on both aspects as illustrated in Figure 6.1. Conversely, behavior is also able to reflect the resident's health to some degree and contribute to the diagnosis. Considering the assessment results, the *m.val* represent the physical ability, and the VAS may have influence on the mental health. For subject-1, for example, the *m.val* indicates that he or she has more mobility, but the VAS indicates that more pain can be felt. From the *change*, this subject's ability to maintain the lifestyle is relatively lower. The possible reason for this result is that the subject has got more physical ability, but the pain drove him or her to a restless life. However, at least by analyzing these four subjects, it is insufficient to predict the future health status merely leaning upon the behavior profiling.

6.3. Chapter Summary

To sum up, through the discussion of the thesis from two aspects, objectives and questions, an unsupervised paradigm to extract behavior related information from sensor-enhanced living environments have been successfully implemented. Some basic behavior information and behavior pattern can be discovered, and the results have also been evaluated considering the medical assessment tests, and been proved to be able to contribute to the patient's diagnosis.

Chapter 7

Conclusions and Outlook

In the context of pervasive healthcare, methods for dealing with the healthcare of older people in their real-life situations is an open research subject. This thesis has proposed a systematic unsupervised methods to discover useful information from the datasets that are collected via sensor-enhanced living environments. The discovered information is able to profile human behavior. The methods are practical, without the limitations such as data labeling and model training, and adaptive, which can be used for various individuals.

The whole work can be divided into three parts. First of all, a general abstraction of in-home monitoring scenarios was carried out, where the *@home graph* was adopted to map the topology and the state change of the components in the graph was used to denote the events occurring in living environment, and the time-line was abstracted through a hierarchical time structure. Secondly, systematic unsupervised methods have been proposed based on *apriori* algorithm. The main procedure of information discovery can be summarized as event extraction, data reorganization, and behavior pattern discovery. On the basis of discovered information, a number of metrics have been defined for profiling human behaviors. Finally, for evaluating the effectiveness of the methods proposed in the current work, the methods and theories are applied to the datasets collected from the GAL-NATARS study, in which there are both sensor data and results of medical assessments from the living environments of real older patients.

In general, through the discussion from two aspects, consisting of the objectives and the questions, it is evident that this thesis has successfully implemented an unsupervised paradigm of extracting behavior related information from sensor-enhanced living environments. The results of applying the methods to the datasets of GAL-NATARS study suggests the objectives have been achieved and the questions have also been answered. The basic behavior information and the behavior pattern can be discovered. The results of the methods of this thesis have been investigated by analyzing the correlation to the assessment results, which is the benchmark of the subject's health status, it is proved that the methods are able to provide valuable behavior related information for fully inspecting the patients' health, even though the dominant relationship between the behavior profile and the health status haven't shown yet.

By investigating the results of the current work, it is recommended that future work on this topic be under taken in the following areas: First, a large scale study is in demand. A study with more subjects and longer study duration would be helpful for formulating more general model to describe the correlation between human behavior and health status. To investigate the performance of the methods, a comparative study between the subjects with different health condition would be helpful. Second, more appropriate sensors should be considered. Through investigating the results, some shortages of the sensors used in the GAL-NATARS study has been exposed. For particular purposes, the sensor with higher sensitivity and accuracy should be taken into account, such as light barrier for passing through a particular zone. Third, it is necessary to address the issue of more than one residents. The current work is based on the assumption that the resident lives alone. In the situation of more than one resident, probably it is possible to use other approaches to distinguish data from different residents.

In conclusion, this thesis has successfully proposed systematical unsupervised methods for exploring the application of sensor-enhanced living environments. The results of applying the methods to the datasets of a study have been proved to be valuable for health assessments. The research will serve as a base for future studies of information discovery in the setting of sensor-enhanced environment.

Bibliography

[1] Kinsella K, He W. An aging world: 2008. International population report. Washington, DC: U.S. Census Bureau; 2009.

[2] Koch S, Marschollek M, Wolf KH, Plischke M, Haux R. On health-enabling and ambient-assistive technologies. What has been achieved and where do we have to go? Methods Inf Med. 2009;48(1):29–37.

[3] Department of Economic and Social Affairs and Population Division. World Population Ageing 2009. New York: United Nations; 2009.

[4] Weiser M. The computer for the 21st century. SIGMOBILE Mob Comput Commun Rev. 1999 July;3(3):3–11.

[5] Cook DJ, Das SK. Pervasive computing at scale: Transforming the state of the art. Pervasive and Mobile Computing. 2012;8(1):22–35.

[6] Korhonen I, Bardram JE. Guest Editorial Introduction to the Special Section on Pervasive Healthcare. Information Technology in Biomedicine, IEEE Transactions on. 2004 sept;8(3):229–234.

[7] Orwat C, Graefe A, Faulwasser T. Towards pervasive computing in health care - A literature review. BMC Medical Informatics and Decision Making. 2008;8(1):26.

[8] Varshney U. Pervasive healthcare. Computer. 2003 dec;36(12):138–140.

[9] Arnrich B, Mayora O, Bardram J, Troster G. Pervasive healthcare: paving the way for a pervasive, user-centered and preventive healthcare model. Methods Inf Med. 2010;49(1):67–73.

[10] Dishman E. Inventing Wellness Systems for Aging in Place. Computer. 2004 May;37(5):34–41.

[11] Korhonen I, Parkka J, Van Gils M. Health monitoring in the home of the future. Engineering in Medicine and Biology Magazine, IEEE. 2003 may-june;22(3):66–73.

[12] Intille SS, Larson K, Tapia EM, Beaudin JS, Kaushik P, Nawyn J, et al. Using a live-in laboratory for ubiquitous computing research. In: Proceedings of the 4th international conference on Pervasive Computing. PERVASIVE'06. Berlin, Heidelberg: Springer-Verlag; 2006. p. 349–365.

[13] Rantz M, Aud M, Alexander G, Oliver D, Minner D, Skubic M, et al. Tiger-Place: An Innovative Educational and Research Environment. In: AAAI Fall Symposium; 2008. p. 84–91.

[14] CASAS; 2012. Accessed: 23/02/2013. `http://ailab.eecs.wsu.edu/casas/`.

[15] Rashidi P, Cook DJ. Keeping the Resident in the Loop: Adapting the Smart Home to the User. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on. 2009 sept;39(5):949–959.

[16] Aware Home Research Initiative (AHRI); 2012. Accessed: 23/02/2013. Available from: `http://awarehome.imtc.gatech.edu/`.

[17] Abowd GD, Mynatt ED. In: Cook DJ, Das SK, editors. Designing for the Human Experience in Smart Environments. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc.; 2005. p. 151–174.

[18] van Halteren A, Bults R, Wac K, Dokovsky N, Koprinkov G, Widya I, et al. Wireless body area networks for healthcare: the MobiHealth project. Studies in health technology and informatics. 2004;108:181–93.

[19] Gietzelt M, Nemitz G, Wolf KH, Meyer Zu Schwabedissen H, Haux R, Marschollek M. A clinical study to assess fall risk using a single waist accelerometer. Informatics for Health and Social Care. 2009;34(4):181–188.

[20] Lau Hy, Tong Ky, Zhu H. Support vector machine for classification of walking conditions of persons after stroke with dropped foot. Human Movement Science. 2009;28(4):504–514.

[21] Alaqtash M, Yu H, Brower R, Abdelgawad A, Sarkodie-Gyan T. Application of wearable sensors for human gait analysis using fuzzy computational algorithm. Engineering Applications of Artificial Intelligence. 2011;24(6):1018–1025.

[22] Atallah L, Yang GZ. The use of pervasive sensing for behaviour profiling? a survey. Pervasive and Mobile Computing. 2009;5(5):447–464.

[23] Wray RE, Laird JE. Variability in human behavior modeling for military simulations. In: In Proceedings of Behavior Representation in Modeling and Simulation Conference (BRIMS); 2003. .

[24] Chikhaoui B, Wang S, Pigot H. ADR-SPLDA: Activity discovery and recognition by combining sequential patterns and latent Dirichlet allocation. Pervasive and Mobile Computing. 2012;8(6):845–862.

[25] Duda RO, Hart PE, Stork DG. Pattern Classification (2nd Edition). Wiley-Interscience; 2000.

[26] Virone G. Assessing everyday life behavioral rhythms for the older generation. Pervasive and Mobile Computing. 2009;5(5):606–622.

[27] Noury N, Quach KA, Berenguer M, Teyssier H, Bouzid MJ, Goldstein L, et al. Remote follow up of health through the monitoring of electrical activities on the residential power line: preliminary results of an experimentation. In: Proceedings of the 11th international conference on e-Health networking, applications and services. Healthcom'09. Piscataway, NJ, USA: IEEE Press; 2009. p. 9–13.

[28] Karunanithi M. Monitoring technology for the elderly patient. Expert Review of Medical Devices. 2007;4(2):267–277.

[29] Stankovic JA, Cao Q, Doan T, Fang L, He Z, Kiran R, et al. Wireless Sensor Networks for In-Home Healthcare: Potential and Challenges. In: High Confidence Medical Device Software and Systems (HCMDSS) Workshop. Philadelphia; 2005. .

[30] Alemdar H, Ersoy C. Wireless sensor networks for healthcare: A survey. Computer Networks. 2010;54(15):2688–2710.

[31] Chan M, Campo E, Estève D, Fourniols JY. Smart homes - Current features and future perspectives. Maturitas. 2009;64(2):90–97.

[32] Demiris G. Home based E-health applications. Studies in health technology and informatics. 2004;106:15–24.

[33] World Health Origanization. Telemedicine: Opportunities and developments in member states; 2010. `www.who.int/goe/publications/goe_telemedicine_2010.pdf`.

[34] Chan M, Estève D, Escriba C, Campo E. A review of smart homes – Present state and future challenges. Computer Methods and Programs in Biomedicine. 2008;91(1):55–81.

[35] Haux R, Howe J, Marschollek M, Plischke M, Wolf KH. Health-enabling technologies for pervasive health care: on services and ICT architecture paradigms. Informatics for Health and Social Care. 2008;33(2):77–89.

[36] Bott OJ, Marschollek M, Wolf KH, Haux R. Towards new scopes: sensor-enhanced regional health information systems-part 1: architectural challenges. Methods of information in medicine. 2007;46(4):476–83.

[37] Wolf KH, Marschollek M, Bott OJ, Howe J, Haux R. Sensors for Health-related Parameters and Data Fusion Approaches. In: ECEH; 2007. p. 155–162.

[38] García-Sáez G, Hernando ME, Martínez-Sarriegui I, Rigla M, Torralba V, Brugués E, et al. Architecture of a wireless Personal Assistant for telemedical diabetes care. International Journal of Medical Informatics. 2009;78(6):391–403.

[39] de Lusignan S, Wells S, Johnson P, Meredith K, Leatham E. Compliance and effectiveness of 1 year's home telemonitoring. The report of a pilot study of patients with chronic heart failure. European Journal of Heart Failure. 2001;3(6):723–730.

[40] Yuce MR. Implementation of wireless body area networks for healthcare systems. Sensors and Actuators A: Physical. 2010;162(1):116–129.

[41] Vogel S, Hulsbusch M, Hennig T, Blazek V, Leonhardt S. In-Ear Vital Signs Monitoring Using a Novel Microoptic Reflective Sensor. Information Technology in Biomedicine, IEEE Transactions on. 2009 nov;13(6):882 –889.

[42] Amft O, Troester G. Recognition of dietary activity events using on-body sensors. Artificial Intelligence in Medicine. 2008;42(2):121–136.

[43] Salarian A, Russmann H, Vingerhoets FJG, Burkhard PR, Aminian K. Ambulatory Monitoring of Physical Activities in Patients With Parkinson's Disease. Biomedical Engineering, IEEE Transactions on. 2007 Dec;54(12):2296–2299.

[44] Fleury A, Noury N, Vacher M, Glasson H, Seri JF. Sound and speech detection and classification in a Health Smart Home. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE; 2008. p. 4644–4647.

[45] de Bruin ED, Hartmann A, Uebelhart D, Murer K, Zijlstra W. Wearable systems for monitoring mobility-related activities in older people: a systematic review. Clinical Rehabilitation. 2008 Oct-Nov;22(10-11):878–95.

[46] Huynh T, Fritz M, Schiele B. Discovery of activity patterns using topic models. In: Proceedings of the 10th international conference on Ubiquitous computing. UbiComp '08. New York, NY, USA: ACM; 2008. p. 10–19.

[47] Mathie M. Monitoring and interpreting human movement patterns using a triaxial accelerometer. University of New South Wales; 2003. PhD thesis.

[48] Demiris G, Hensel BK. Technologies for an aging society: a systematic review of "smart home" applications. Yearbook of medical informatics. 2008;p. 33–40.

[49] Gietzelt M, Wolf KH, Haux R. A nomenclature for the analysis of continuous sensor and other data in the context of health-enabling technologies. Studies in health technology and informatics. 2011;169:460–4.

[50] Bellazzi R, Zupan B. Intelligent Data Analysis ? Special Issue. Methods of Information in Medicine. 2001;40(5):362–364.

[51] Bellazzi R, Diomidous M, Sarkar IN, Takabayashi K, Ziegler A, McCray AT. Data analysis and data mining: current issues in biomedical informatics. Methods Inf Med. 2011;50(6):536–44.

[52] Haux R. Individualization, globalization and health ? about sustainable information technologies and the aim of medical informatics. International Journal of Medical Informatics. 2006;75(12):795–808.

[53] Peek N, Swift S. Intelligent data analysis for knowledge discovery, patient monitoring and quality assessment. Methods Inf Med. 2012 Aug;51(4):318–22.

[54] Preece SJ, Goulermas JY, Kenney LPJ, Howard D, Meijer K, Crompton R. Activity identification using body-mounted sensors?a review of classification techniques. Physiological Measurement. 2009;30(4):R1.

[55] Webb AR. Statistical Pattern Recognition. 2nd ed. Chichester, West Sussex PO19 8SQ, England: John Wiley & Sons; 2002.

[56] Sharma N, Gedeon T. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. Comput Methods Prog Biomed. 2012 Dec;108(3):1287–1301.

[57] Ye J, Dobson S, McKeever S. Situation identification techniques in pervasive computing: A review. Pervasive Mob Comput. 2012 Feb;8(1):36–66.

[58] Nyan MN, Tay FEH, Tan AWY, Seah KHW. Distinguishing fall activities from normal activities by angular rate characteristics and high-speed camera characterization. Medical Engineering & Physics. 2006;28(8):842–849.

[59] Bishop CM. Neural Networks for Pattern Recognition. New York, NY, USA: Oxford University Press, Inc.; 1995.

[60] Shahar Y. Timing is Everything: Temporal Reasoning and Temporal Data Maintenance in Medicine. In: in Medicine, Lecture Notes In Artificial Intelligence 1620. Springer Verlag; 1999. p. 30–46.

[61] Roddick JF, Society IC, Spiliopoulou M, Society IC. A Survey of Temporal Knowledge Discovery Paradigms and Methods. IEEE Transactions on Knowledge and Data Engineering. 2002;14:750–767.

[62] Antunes CM, Oliveira AL. Temporal Data Mining : an overview. Lecture Notes in Computer Science. 2001;p. 1–15.

[63] Chiu B, Keogh E, Lonardi S. Probabilistic discovery of time series motifs. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. KDD '03. New York, NY, USA: ACM; 2003. p. 493–498.

[64] Chikhaoui B, Wang S, Pigot H. ADR-SPLDA: Activity discovery and recognition by combining sequential patterns and latent Dirichlet allocation. Pervasive and Mobile Computing. 2012 Dec;8(6):845–862.

[65] Duchêne F, Garbay C, Rialle V. Learning recurrent behaviors from heterogeneous multivariate time-series. Artificial Intelligence in Medicine. 2007;39(1):25–47.

[66] Jakkula V, Cook DJ. Anomaly detection using temporal data mining in a smart home environment. Methods Inf Med. 2008;47(1):70–5.

[67] Stenvall M, Olofsson B, Nyberg L, Lundstrom M, Gustafson Y. Improved performance in activities of daily living and mobility after a multidisciplinary postoperative rehabilitation in older people with femoral neck fracture: a randomized controlled trial with 1-year follow-up. J Rehabil Med. 2007;39(3):232–8.

[68] Blocker WPJ. Maintaining functional independence by mobilizing the aged. Geriatrics. 1992;47(1):42, 48–50, 53 passim.

[69] Bharucha AJ, Pandav R, Shen C, Dodge HH, Ganguli M. Predictors of Nursing Facility Admission: A 12-Year Epidemiological Study in the United States. Journal of the American Geriatrics Society. 2004;52(3):434–439.

[70] McCallum J, Simons LA, Simons J, Friedlander Y. Patterns and predictors of nursing home placement over 14 years: Dubbo study of elderly Australians. Australasian Journal on Ageing. 2005;24(3):169–173.

[71] Mehta KM, Yaffe K, Covinsky KE. Cognitive Impairment, Depressive Symptoms, and Functional Decline in Older People. Journal of the American Geriatrics Society. 2002;50(6):1045–1050.

[72] Katz S, Downs TD, Cash HR, Grotz RC. Progress in Development of the Index of ADL. The Gerontologist. 1970;10(1 Part 1):20–30.

[73] Lawton MP, Brody EM. Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. The Gerontologist. 1969;9(3 Part 1):179–186.

[74] Monk TH, Reynolds CFr, Kupfer DJ, Hoch CC, Carrier J, Houck PR. Differences over the life span in daily life-style regularity. Chronobiology international. 1997 May;14(3):295–306.

[75] Duffy JF, Dijk DJ, Hall EF, Czeisler CA. Relationship of endogenous circadian melatonin and temperature rhythms to self-reported preference for morning or evening activity in young and older people. J Investig Med. 1999;47(3):141–50.

[76] Monk TH, Flaherty JF, Frank E, Hoskinson K, Kupfer DJ. The Social Rhythm Metric An Instrument to Quantify the Daily Rhythms of Life. The Journal of Nervous and Mental Disease. 1990;178(2):120–6.

[77] Prigerson HG, Monk TH, Reynolds CF, Begley A, Houck PR, Bierhals AJ, et al. Lifestyle regularity and activity level as protective factors against bereavement-related depression in late-life. Depression. 1995;3(6):297–302.

[78] Camara Magalhaes S, Vitorino Souza C, Rocha Dias T, Felipe Carvalhedo de Bruin P, de Bruin VMS. Lifestyle regularity measured by the social rhythm metric in Parkinson's disease. Chronobiol Int. 2005;22(5):917–24.

[79] McHorney CA. Health status assessment methods for adults: past accomplishments and future challenges. Annu Rev Public Health. 1999;20:309–35.

[80] Torrance GW. Utility approach to measuring health-related quality of life. Journal of Chronic Diseases. 1987;40(6):593–600.

[81] Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. The Journal of rheumatology. 2000;27(7):1745–52.

[82] Natterlund B, Ahlstrom G. Activities of daily living and quality of life in persons with muscular dystrophy. J Rehabil Med. 2001 Sep;27(7):206–11.

[83] Juenger J, Schellberg D, Kraemer S, Haunstetter A, Zugck C, Herzog W, et al. Health related quality of life in patients with congestive heart failure: comparison with other chronic diseases and relation to functional variables. Heart. 2002;87(3):235–241.

[84] Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JB, deGruy FVr, et al. Health-related quality of life in primary care patients with mental disorders: Results from the prime-md 1000 study. JAMA. 1995;274(19):1511–1517.

[85] Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. BMJ (Clinical research ed). 1992 Jul;305(6846):160–4.

[86] Furlong WJ, Feeny DH, Torrance GW, Barr RD. The Health Utilities Index (HUI®) system for assessing health-related quality of life in clinical studies. Annals of Medicine. 2001 Jul;33(5):375–384.

[87] Nelson D, Holtzman D, Bolen J, Stanwyck C, Mack K. Reliability and validity of measures from the Behavioral Risk Factor Surveillance System (BRFSS). Sozial- und Praventivmedizin. 2001;46 Suppl 1:S3–42.

[88] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. SIGMOD Rec. 1993 Jun;22(2):207–216.

[89] Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. In: Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy R, editors. Advances in Knowledge Discovery and Data Mining. Menlo Park, CA, USA: American Association for Artificial Intelligence; 1996. p. 307–328.

[90] Papamatthaiakis G, Polyzos GC, Xylomenos G. Monitoring and modeling simple everyday activities of the elderly at home. In: Proceedings of the 7th IEEE conference on Consumer communications and networking conference. CCNC'10. Piscataway, NJ, USA: IEEE Press; 2010. p. 617–621.

[91] Nazerfard E, Rashidi P, Cook DJ. Discovering Temporal Features and Relations of Activity Patterns. In: Data Mining Workshops (ICDMW), 2010 IEEE International Conference on; 2010. p. 1069–1075.

[92] Jakkula V, Cook DJ, Crandall AS. Temporal pattern discovery for anomaly detection in a smart home. In: Intelligent Environments, 2007. IE 07. 3rd IET International Conference on; 2007. p. 339–345.

[93] Borgelt C, Kruse R. Induction of Association Rules: Apriori Implementation. In: Härdle W, Rönz B, editors. Compstat. Physica-Verlag HD; 2002. p. 395–400. Available from: `http://dx.doi.org/10.1007/978-3-642-57489-4_59`.

[94] Apel M, Bauer J, Haux R, Kolb G, Marschollek M. Prolektive Beobachtungsstudie zur Nutzung von assistierenden Technologien für das Assessment von Risikoprofilen bei Patienten mit stattgehabter Schenkelhalsfraktur; 2012. Unpublished study plan.

[95] Kolb GF. Medizin des alternden Menschen: Lehrbuch zum Gegenstandskatalog der neuen ÄAppO. Wissenschaftliche; 2009.

[96] Neyer FJ, Felber J, Gebhardt C. Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft. Development and validation of a brief measure of technology commitment. Diagnostica. 2012;58(2):87.

[97] Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, Weerdmeester B, Thomas A, Mclelland IL, editors. Usability evaluation in industry. London: Taylor and Francis; 1996. .

[98] PIR motion sensors; 2012. Accessed: 25/02/2013. `http://www.ladyada.net/learn/sensors/pir.html`.

[99] ELV Elektronik AG. 2-Kanal-Bewegungsmelder für Außeneinsatz FS20 PIRA – Bedinungsanleitung. Leer, Germany; 2009.

[100] Gurevich V. Electric Relays: Principles and Applications. Electrical and Computer Engineering Series. Taylor & Francis; 2010.

[101] ELV Elektronik AG. Erschütterungssensor. ELV Elektronik AG, Leer;.

[102] Allen JF. Maintaining knowledge about temporal intervals. Commun ACM. 1983 Nov;26(11):832–843.

[103] Wickham H. ggplot2: elegant graphics for data analysis. Springer New York; 2009. Available from: `http://had.co.nz/ggplot2/book`.

[104] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013. Available from: `http://www.R-project.org/`.

A. Results of Evident IES Distribution

A.1. Results of Subject-1

See Figure A.1, Figure A.2, and Figure A.3.

A.2. Results of Subject-2

See Figure A.4, Figure A.5, Figure A.6, and Figure A.7.

A.3. Results of Subject-3

See Figure A.8, Figure A.9, and Figure A.10.

A.4. Results of Subject-4

See Figure A.11, Figure A.12, and Figure A.13.

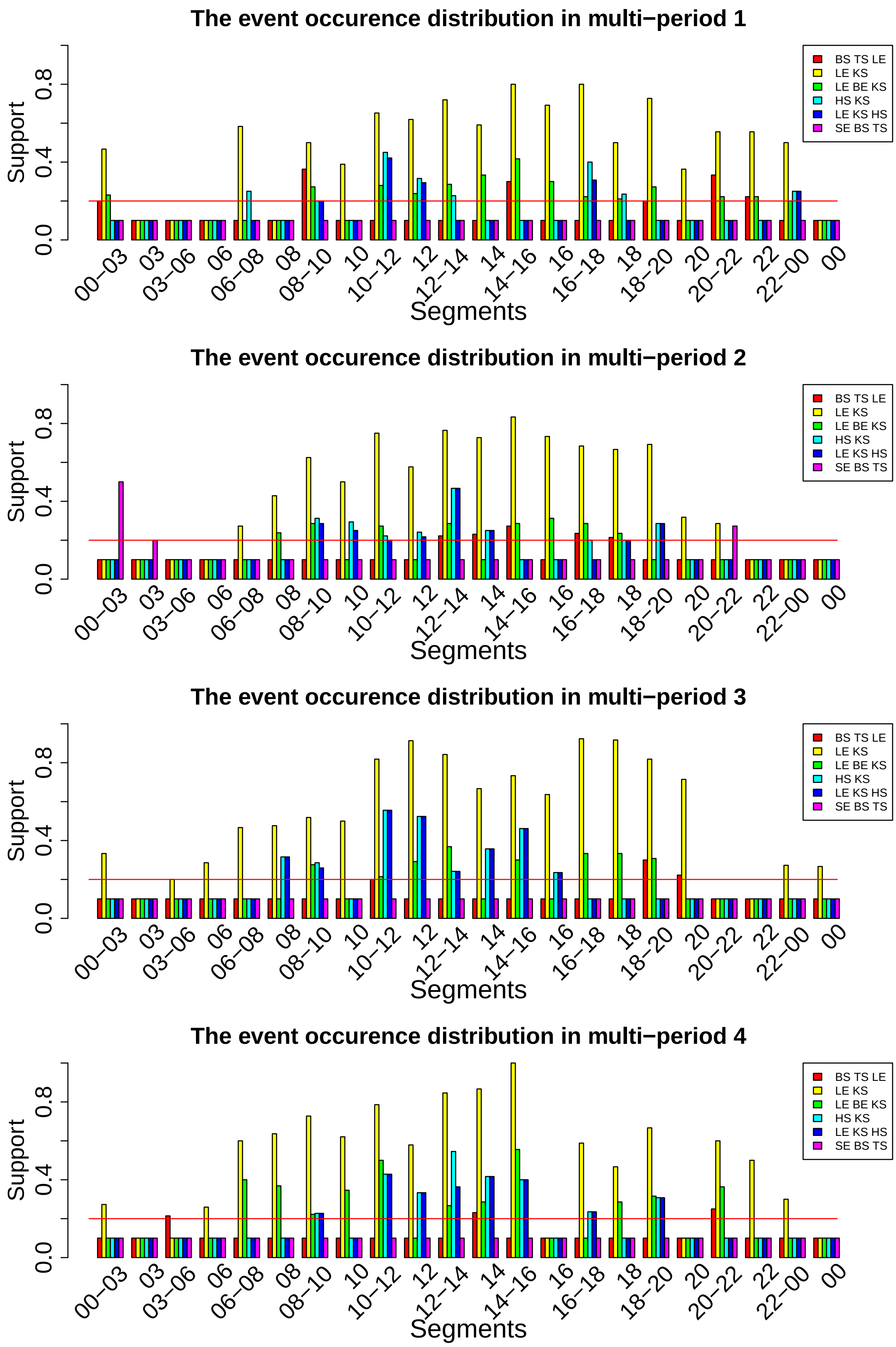


Figure A.1.: The results of subject-1's behavior pattern.

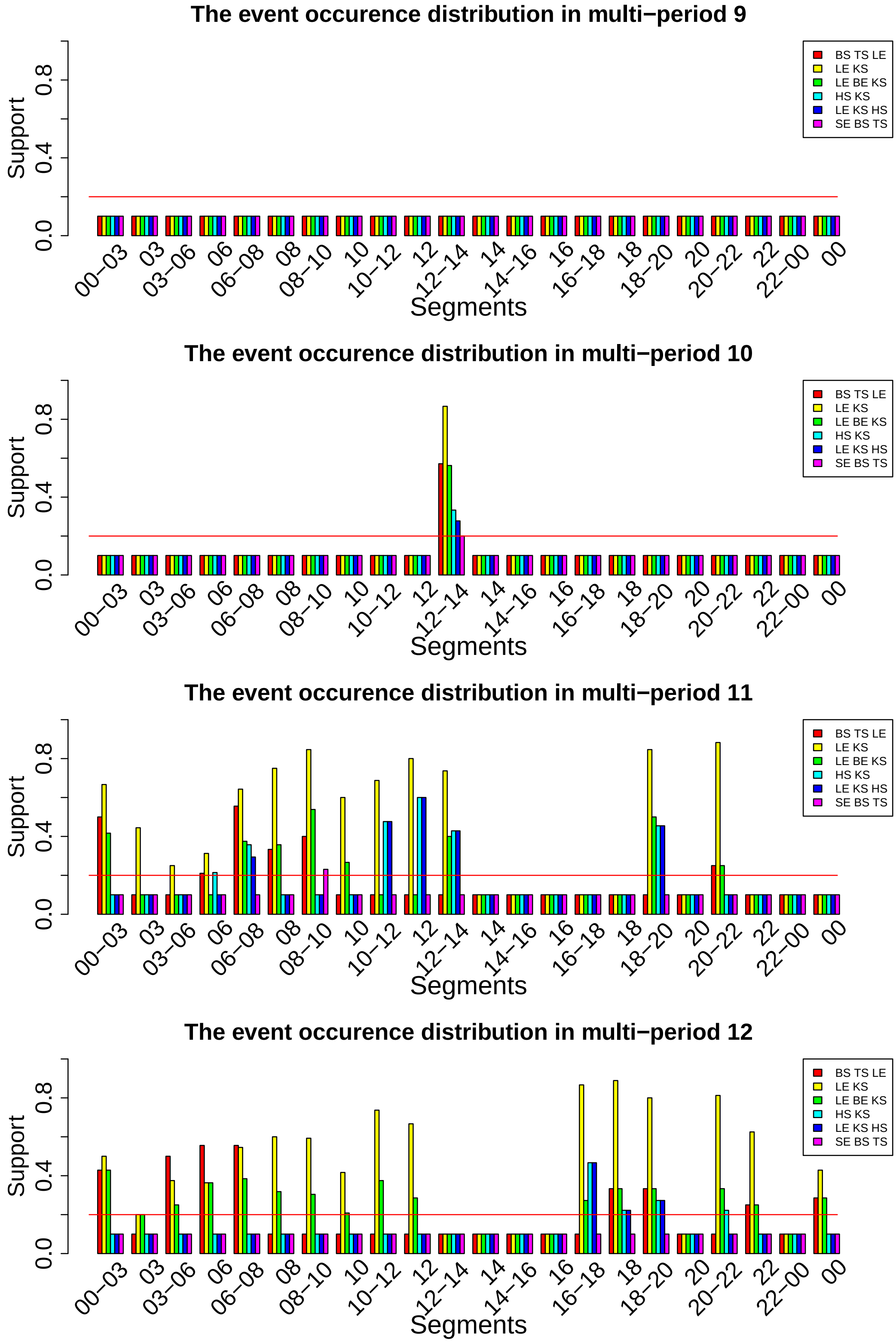


Figure A.2.: The results of subject-1's behavior pattern (continue).

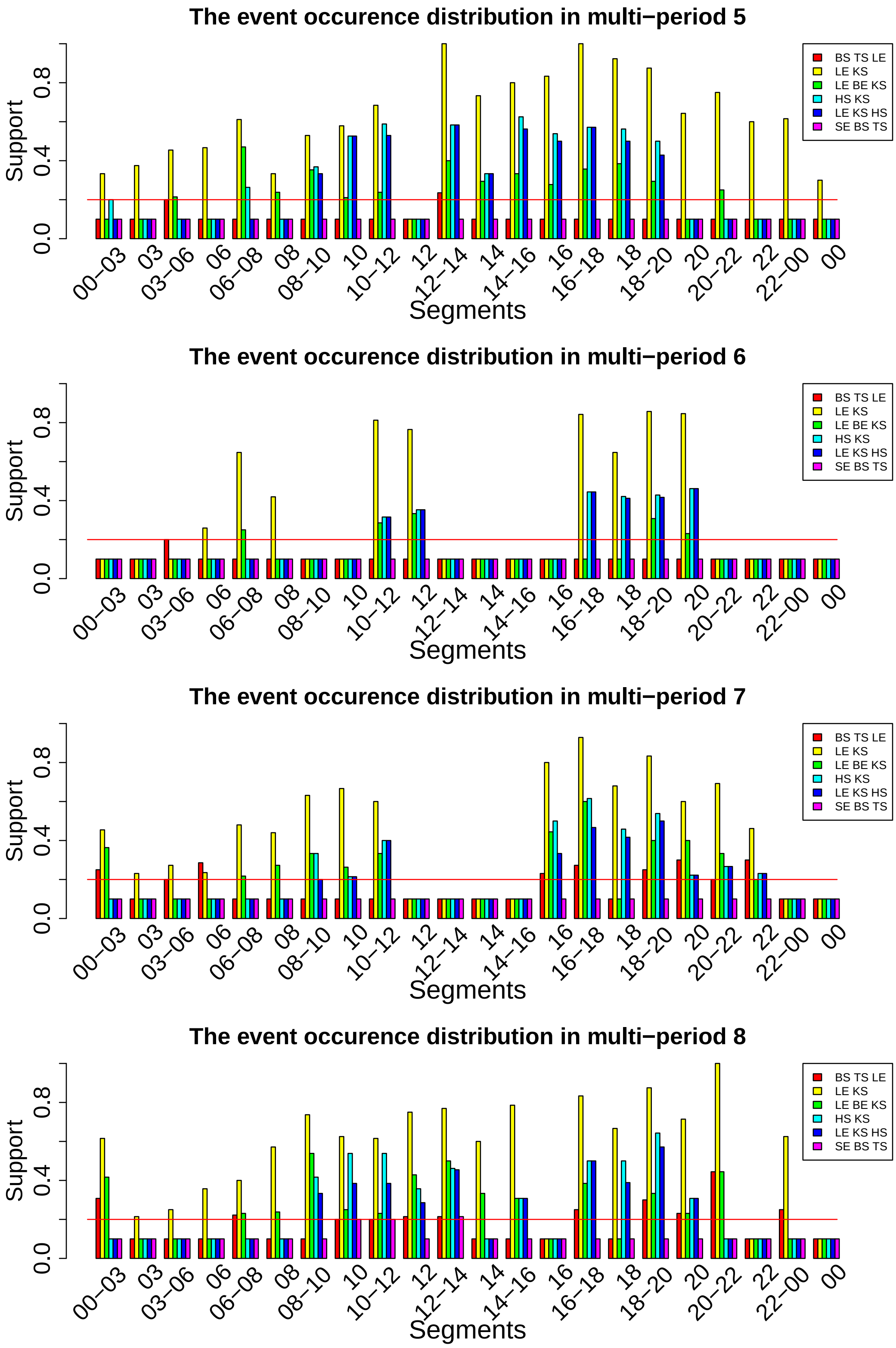


Figure A.3.: **The results of subject-1's behavior pattern (continue).**

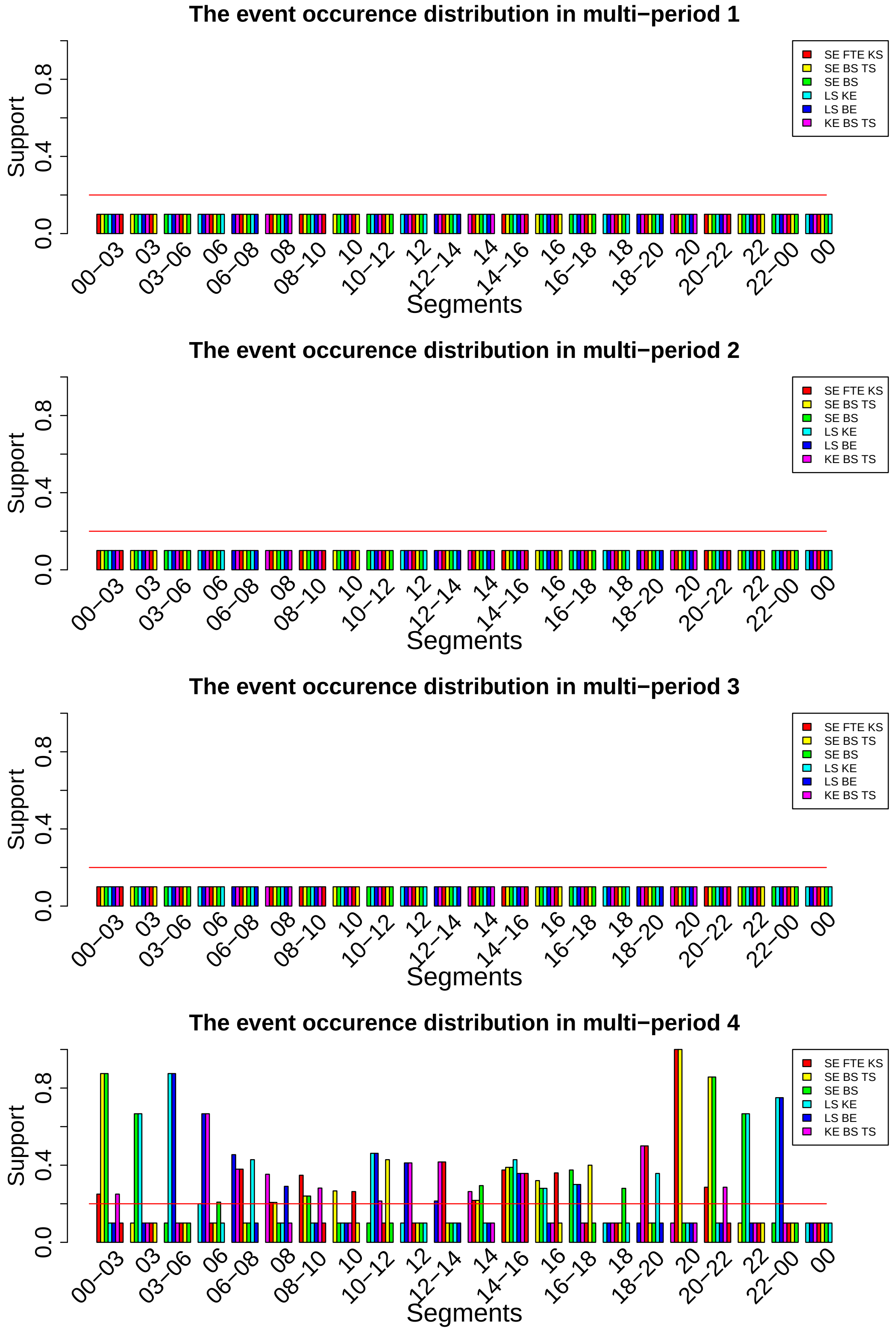


Figure A.4.: **The results of subject-2's behavior pattern.**

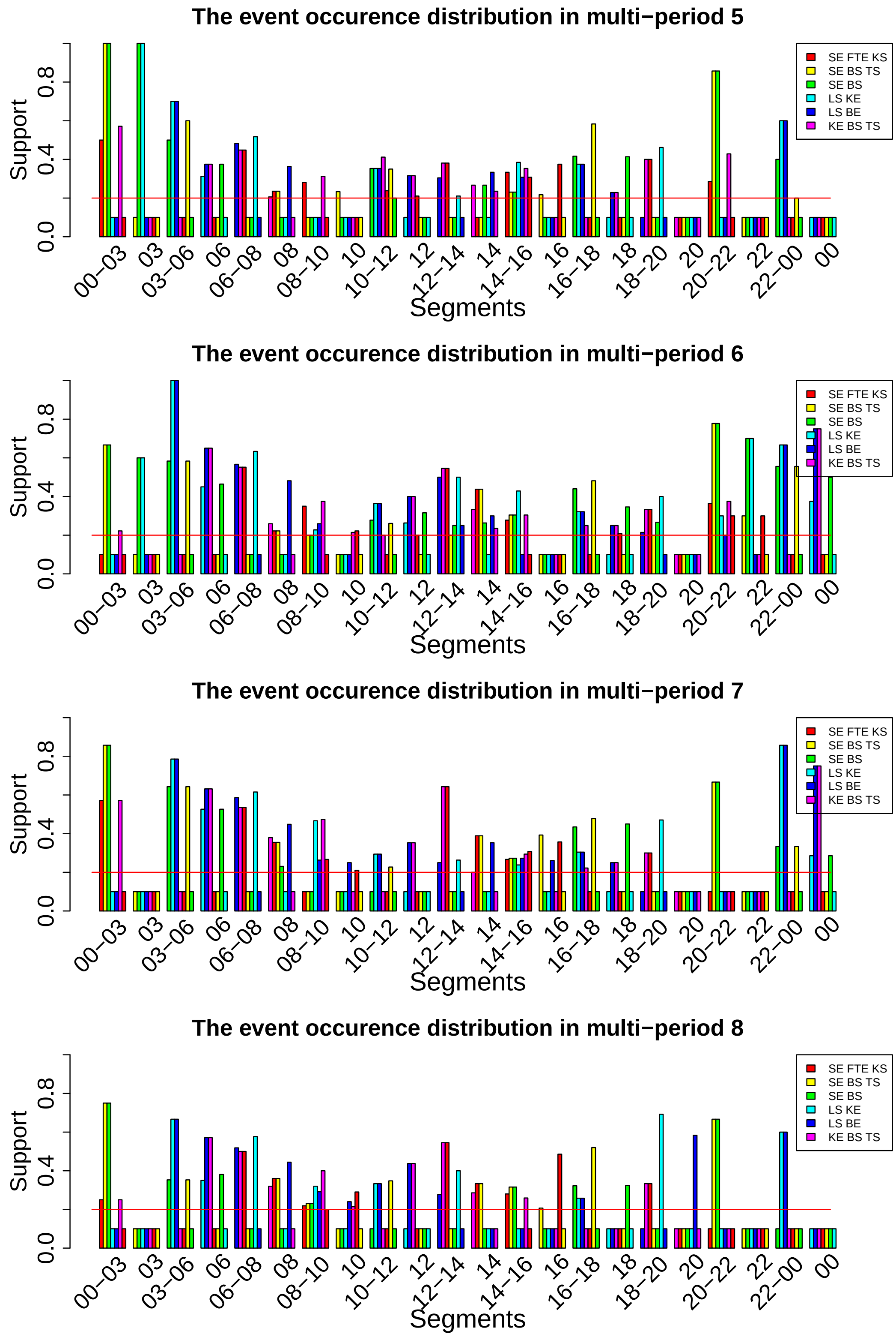


Figure A.5.: **The results of subject-2's behavior pattern (continue).**

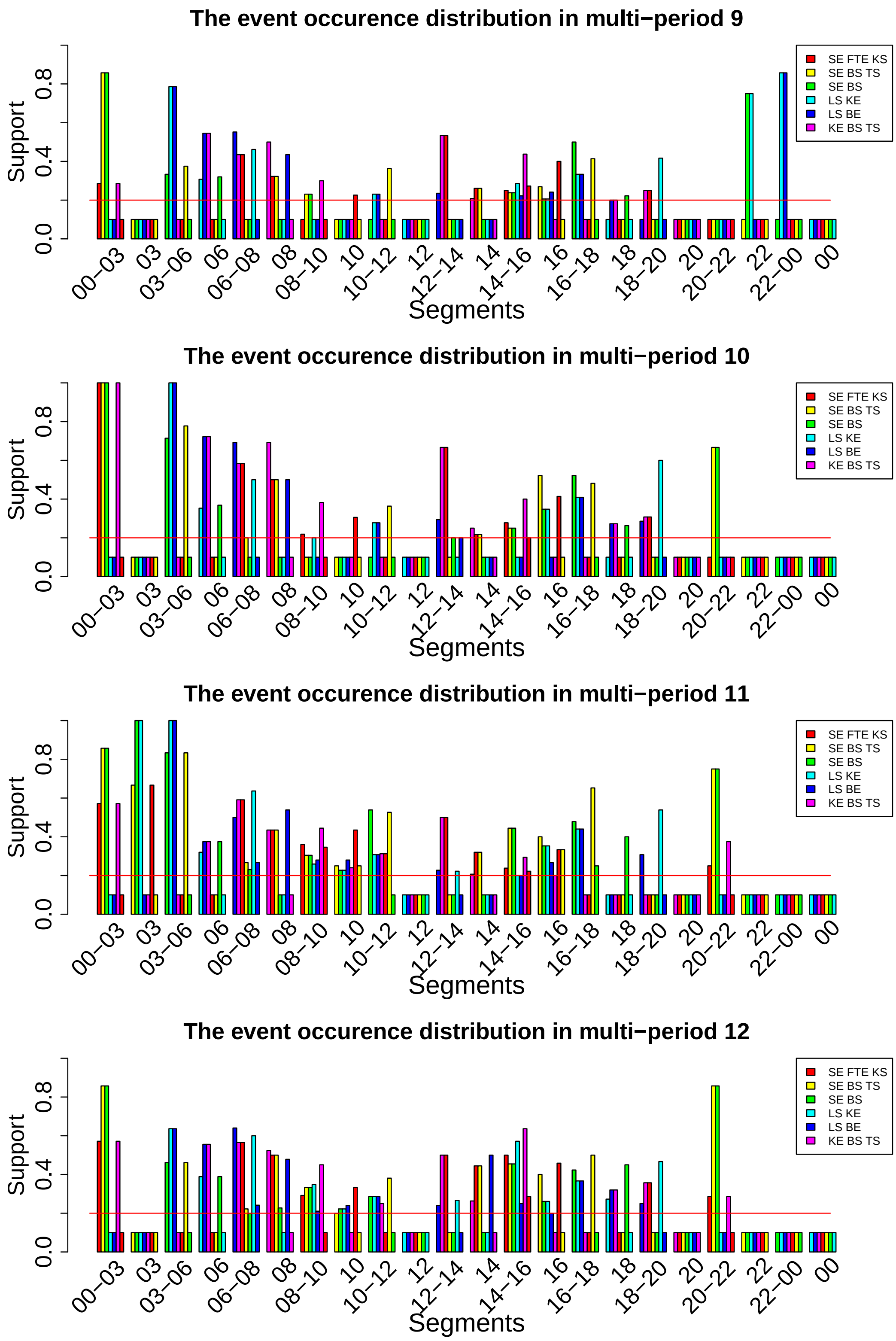


Figure A.6.: **The results of subject-2's behavior pattern (continue).**

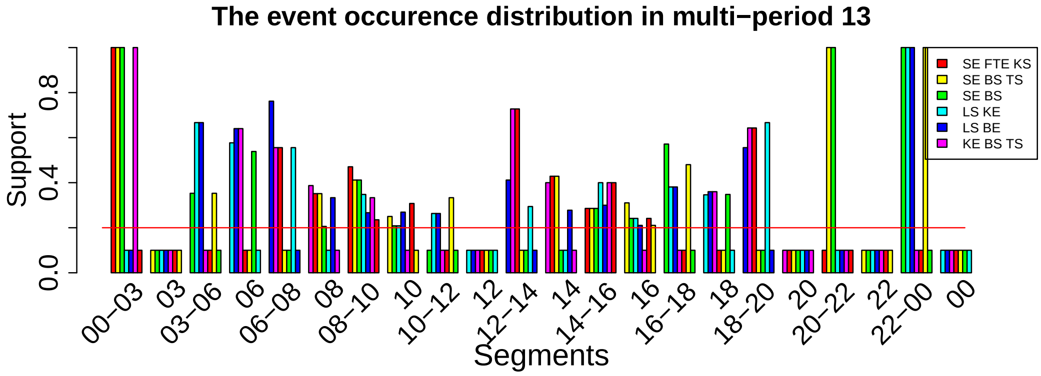


Figure A.7.: The results of subject-2's behavior pattern (continue).

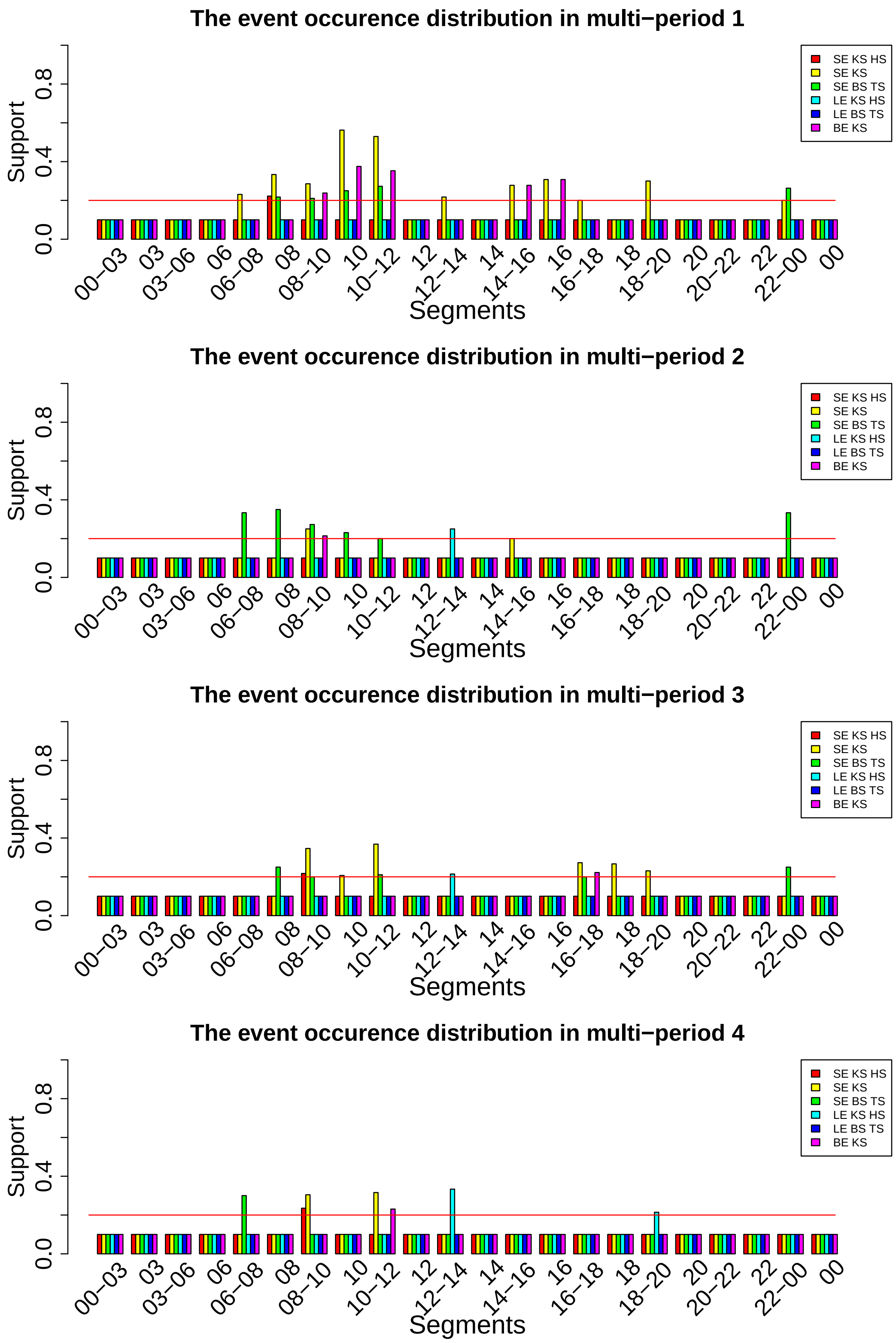


Figure A.8.: **The results of subject-3's behavior pattern.**

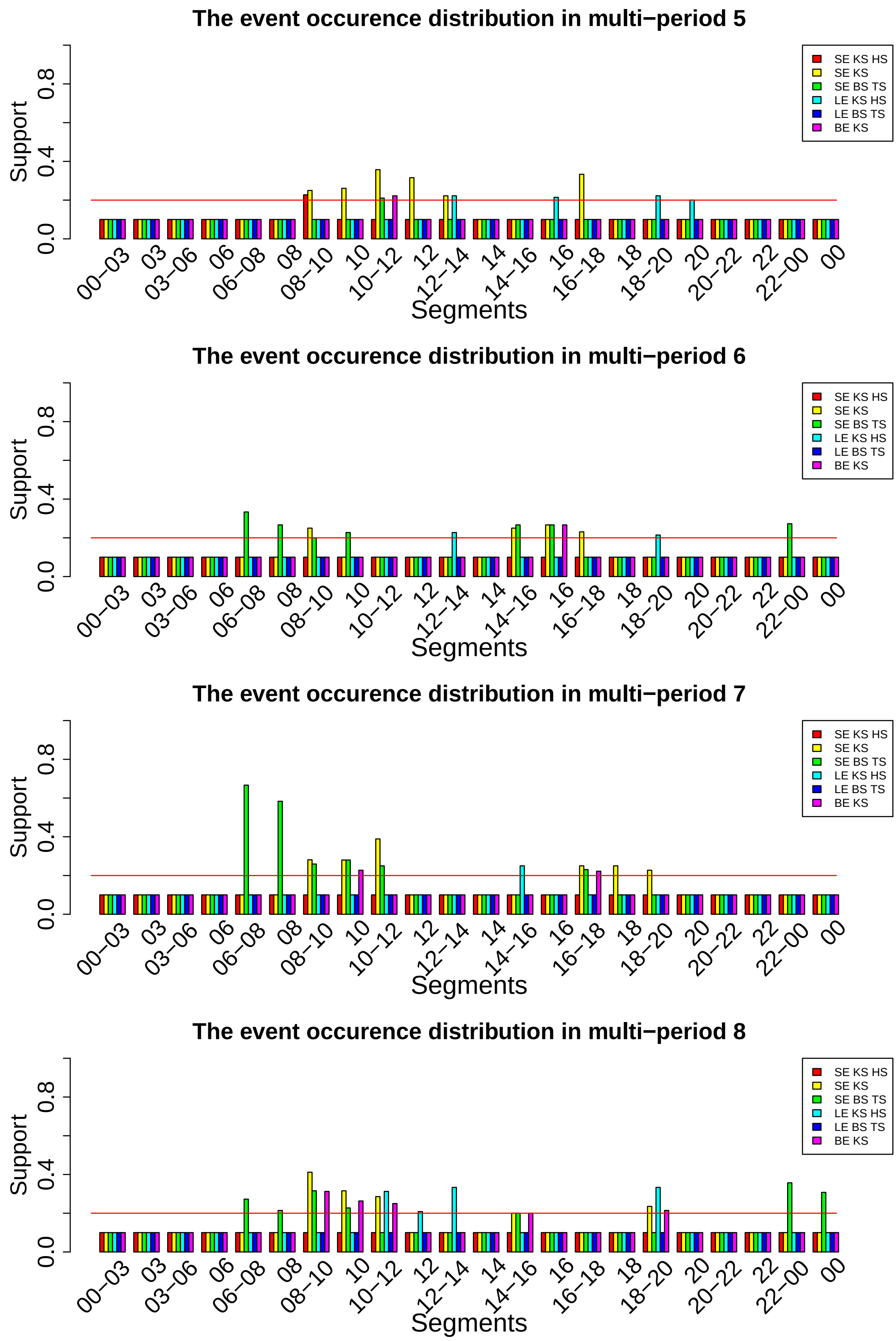


Figure A.9.: **The results of subject-3's behavior pattern (continue).**

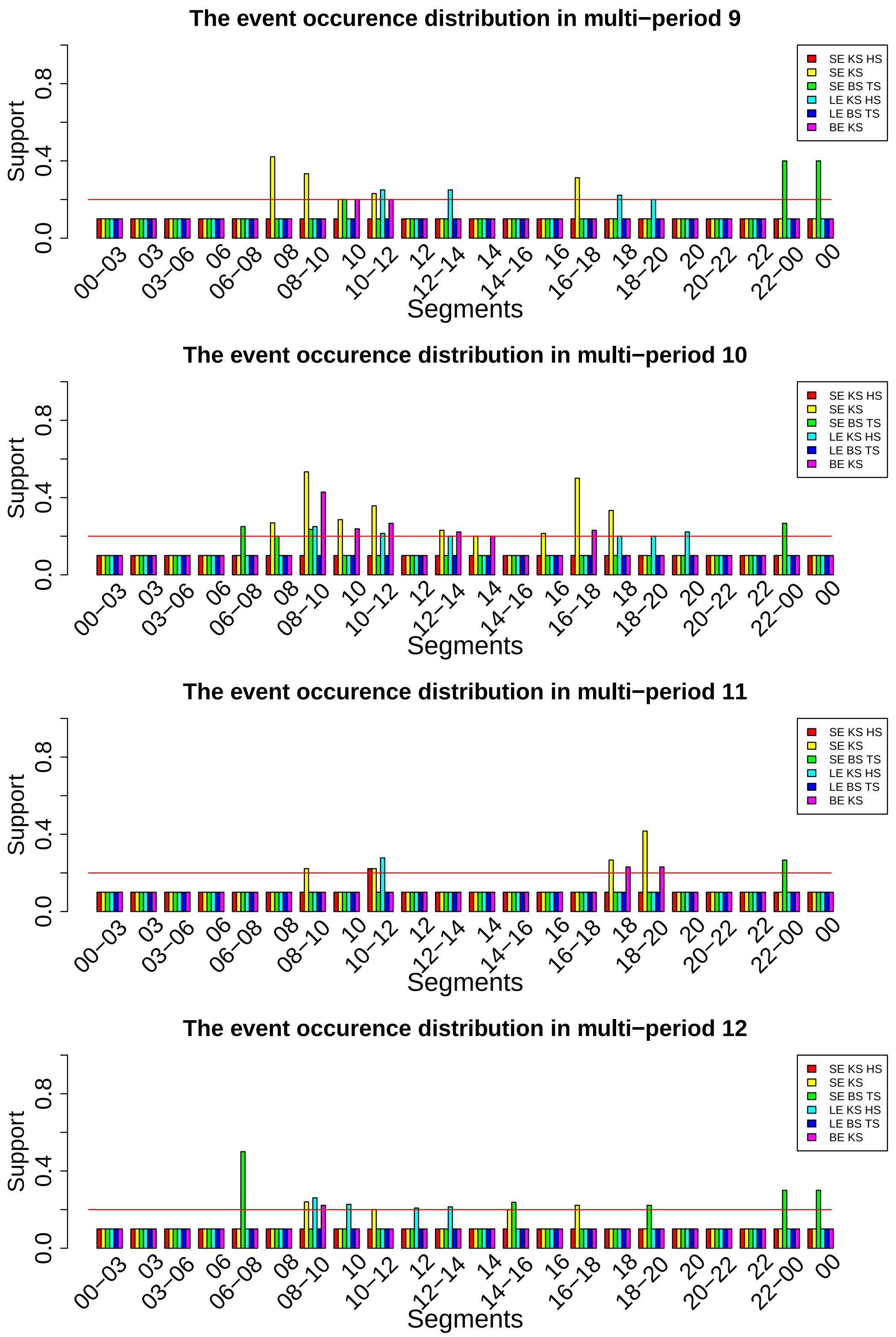


Figure A.10.: **The results of subject-3's behavior pattern (continue).**

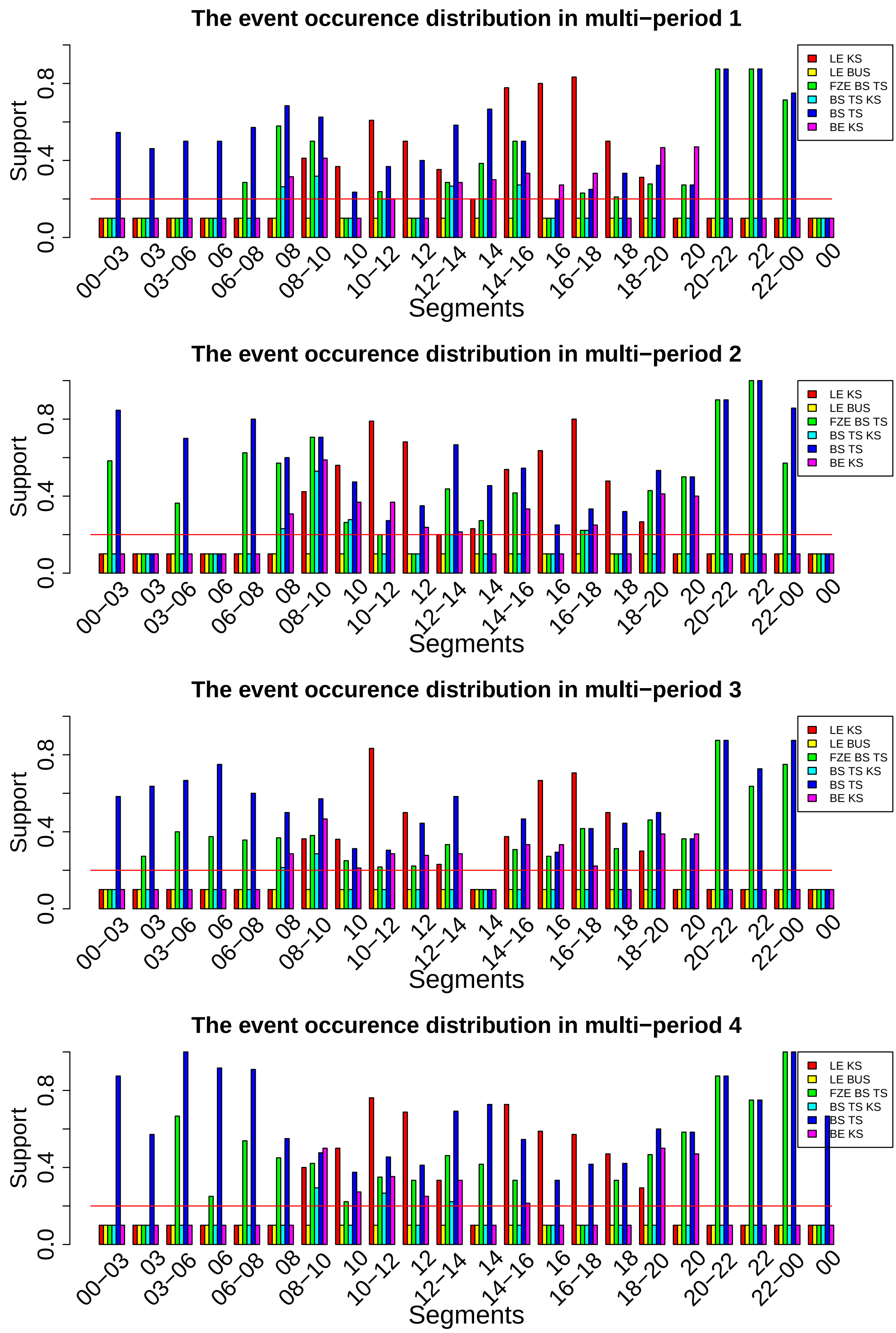


Figure A.11.: The results of subject-4's behavior pattern.

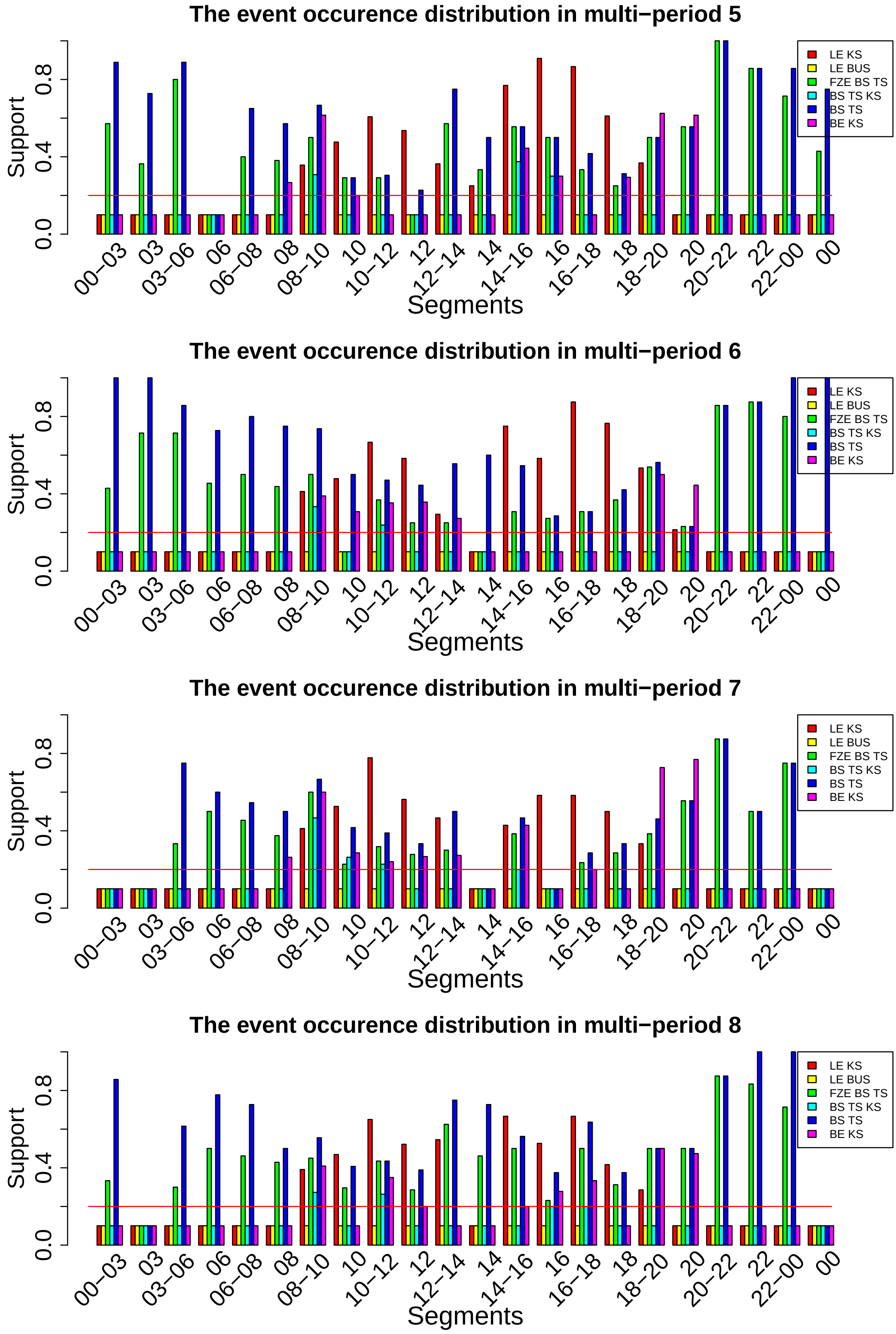


Figure A.12.: The results of subject-4's behavior pattern (continue).

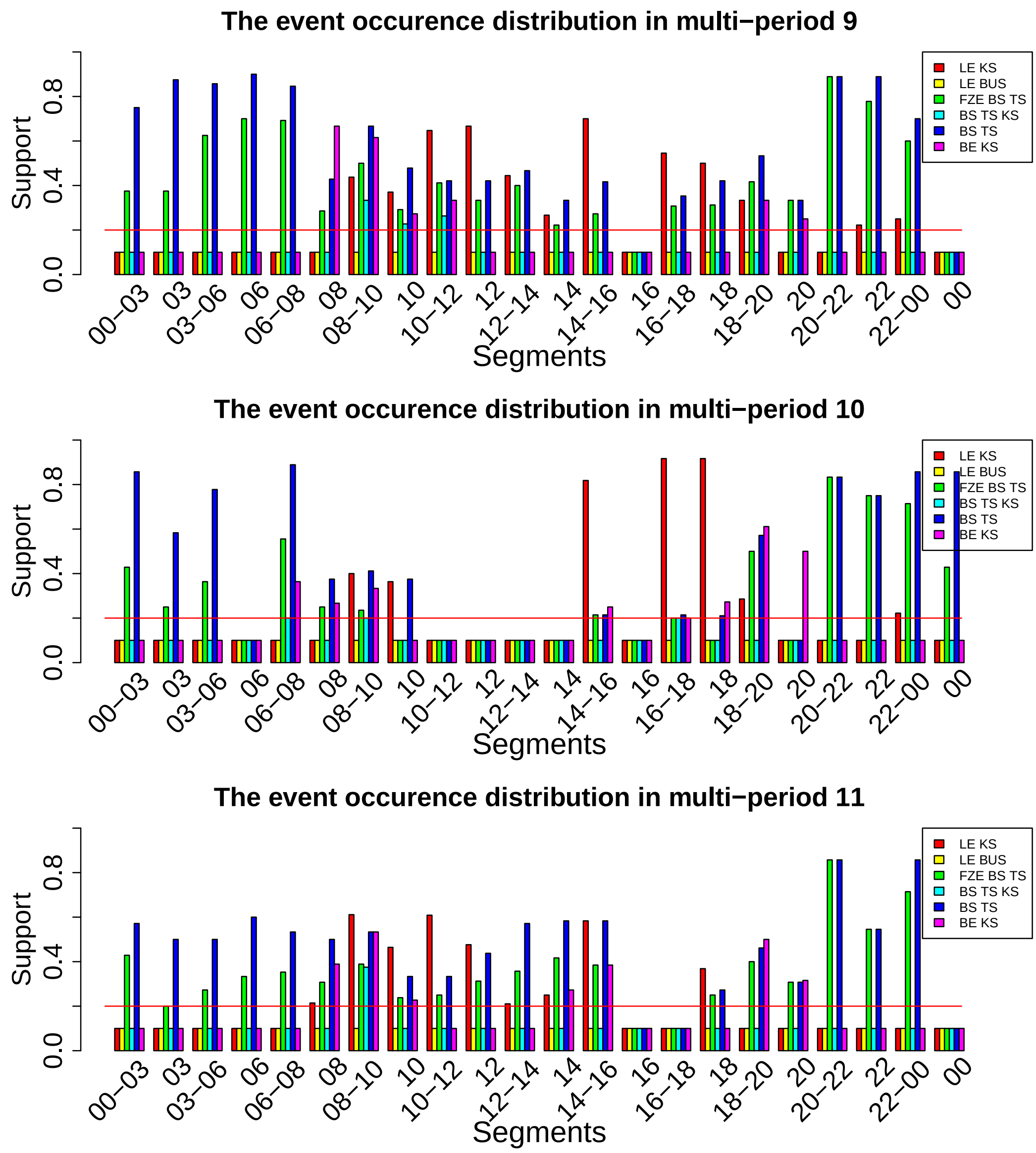


Figure A.13.: The results of subject-4's behavior pattern (continue).

B. The Subjects' Medical Assessments

Table B.1.: Medical assessments of subject-1

AssessmentItems	1	2	3	4	5	6	7	8	9	10	11	12	13
TUG	12.60				11.56				60				11.45
SPPB	3				6				6				5
Tinetti I & II	24				25				25				27
Barthel Index	85				100				100				100
VAS	8	7.3	6.4	5.4	6.7	11	9.5	7	14.5	12.5	9.9	13	13.1

Table B.2.: Medical assessments of subject-2

AssessmentItems	1	2	3	4	5	6	7	8	9	10	11	12	13	14
TUG	32.76				60				27					18.075
SPPB	1				2				1					3.5
Tinetti I & II	17				20				19					22.5
Barthel Index	75				80				85					85
VAS	6			0	0	0	0	3	3	0	0	0	0	0

Table B.3.: Medical assessments of subject-3

AssessmentItems	1	2	3	4	5	6	7	8	9	10	11	12	13
TUG	43.5				42				49				60
SPPB	1				1				1				1
Tinetti I & II	12				17				21				17
Barthel Index	75				75				75				80
VAS	5	12	7	3	6	3	2	2	7	8	9		11

Table B.4.: Medical assessments of subject-4

AssessmentItems	1	2	3	4	5	6	7	8	9	10	11	12	13
TUG	29				22				26				17
SPPB	1				1				1				2
Tinetti I & II	20				23				25				26
Barthel Index	85				85				85				85
VAS	12	6	12	2	8	9	9	1	4	9		2	2